Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Cybernetics

Actigraphy Data Analysis in Bipolar Disorder Patients

Bachelor's thesis

Václav Hlaváč

Study programme: Open Informatics
Branch of study: Computer and Information Science
Supervisor: Ing. Jakub Schneider

Prague, May 2020

BACHELOR'S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Hlaváč Václav** Personal ID number: **474528**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Open Informatics**

Branch of study: **Computer and Information Science**

II. Bachelor's thesis details

Bachelor's thesis title in English:

Actigraphy Data Analysis in Bipolar Disorder Patients

Bachelor's thesis title in Czech:

Analýza aktigrafických dat pacientů trpících bipolární poruchou

Guidelines:

The goal of the thesis is to analyse differences in activity during different symptomatic episodes in the course of bipolar disorder. To do that the student has to familiarise with analysis of actigraphy signals.
1) Familiarise with symptoms of bipolar disorder – mainly focus on differences in physical activity.
2) Crate as simulator of artificial actigraphy data.
3) Study and implement non-parametric analysis used in actigraphy.
4) Analyse complexity of actigraphic signal using sampling entropy.
5) Test effect of missing data on estimation of parameters using simulated and real data.
6) Build a classifier for patient's state using implemented actigraphy parameters.
7) Validate the results on provided patient data.

Bibliography / sources:

[1] S. H. Jones, D. J. Hare, and K. Evershed, "Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder," Bipolar Disord., vol. 7, no. 2, pp. 176–186, 2005.
[2] B. S. B. Gonçalves, P. R. A. Cavalcanti, G. R. Tavares, T. F. Campos, and J. F. Araujo, "Nonparametric methods in actigraphy: An update," Sleep Sci., vol. 7, no. 3, pp. 158–164, Sep. 2014.
[3] K. Krane-Gartiser, T. E. G. Henriksen, G. Morken, A. Vaaler, and O. B. Fasmer, "Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder," PLoS One, vol. 9, no. 2, 2014.
[4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning The Elements of Statistical Learning. 2017.

Name and workplace of bachelor's thesis supervisor:

Ing. Jakub Schneider, Analysis and Interpretation of Biomedical Data, FEE

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: **07.01.2020** Deadline for bachelor thesis submission: **22.05.2020**

Assignment valid until: **30.09.2021**

Ing. Jakub Schneider	doc. Ing. Tomáš Svoboda, Ph.D.	prof. Mgr. Petr Páta, Ph.D.
Supervisor's signature	Head of department's signature	Dean's signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

. _____
Date of assignment receipt

_____
Student's signature

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

In Prague, May 2020

.....
Václav Hlaváč

Abstract

Bipolar disorder is a mental illness with an episodic progress which causes abrupt mood changes, during which a person can experience either depression or mania. These mood changes are often accompanied by changes in behaviour and in the level of physical activity.

This thesis gives an overview of bipolar disorder and actigraphy, a method used to record activity. It also investigates the differences between motor activities recorded during individual episodes and during euthymia. Two methods of activity analysis are described, nonparametric analysis and sample entropy. These methods are then used for the analysis itself, the results of which are consequently communicated. Based on the performed analysis, classificators are created with the aim to differentiate between the activities recorded during depression, mania and euthymia.

Keywords: bipolar disorder, actigraphy, sample entropy, complexity, classification

Abstrakt

Bipolární porucha je duševní onemocnění s epizodickým průběhem, které způsobuje náhlé změny nálady, při kterých se člověk ocitá buď v depresi, nebo v mánii. Tato změna nálady je často doprovázena i změnou v chování a v míře vykazované fyzické aktivity.

Tato práce uvádí krátký přehled informací o bipolární poruše a aktigrafii, metodě používané k záznamu aktivity. Zkoumá také rozdíly v pohybových aktivitách zaznamenaných v průběhu jednotlivých epizod a ve stavu euthymie. Dvě metody analýzy aktivity jsou představeny, neparametrická analýza aktigrafického záznamu a samplovací entropie. Tyto metody jsou poté použity k vlastní analýze, jejíž výsledky jsou následně popsány. Na základě provedené analýzy jsou sestrojeny klasifikátory rozlišující mezi aktivitami zaznamenanými v průběhu deprese, mánie a euthymie.

Klíčová slova: bipolární porucha, aktigrafie, samplovací entropie, komplexita, klasifikace

Acknowledgements

I would like to thank my supervisor Ing. Jakub Schneider for providing helpful guidance throughout the whole course of writing this thesis. I would also like to thank doc. Ing. Daniel Novák, Ph.D., for valuable input during my work on this project. Finally, I would like to thank my family for the constant support that they show me.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Mood swings are a natural part of everyone's life. Feelings of happiness can arrive only to be superseded by moments of sadness not even days later. Many of the changes in our mood can be attributed to an event in our lives and do not cause notable difficulties. However, in some cases, these mood fluctuations attain such levels of severity that they can heavily disrupt the affected person's life and render him or her functionally impaired. Under such circumstances, an affective disorder can be considered to be at play. Bipolar disorder is one of such disorders.

Bipolar disorder is characterized by recurrent episodes of elevated mood called mania and episodes of hopelessness and sadness called depression. The change in mood during an episode is often accompanied by a significant change in motor activity. It is a chronic disorder, but it can be made more manageable through proper diagnosis and treatment.

The main focus of this bachelor's thesis is to examine the differences in activity during each of the episodes of bipolar disorder. An overview of bipolar disorder is given, including a description of the main characteristics of episode types that can occur during the course of this illness. Actigraphy, a technique used to record physical activity, is introduced along with a method developed to generate artificial actigraphic data.

Furthermore, two techniques that can be used to analyze the recorded motor activity are described, namely the nonparametric analysis of an actigraphic signal and sample entropy. The severity of the error that arises when these methods are applied to records that contain missing values is examined, both using the provided acitivity records of bipolar disorder patients and using the generator of artificial data.

After that, a grid search is conducted in order to find settings of sample entropy that would optimize its ability to differentiate between the individual mood states.

The behaviour of variables studied using nonparametric analysis is then compared between the different episodes of bipolar disorder and with the expected behaviour.

Ultimately, machine learning classification models are applied together with the features extracted using the above-mentioned methods of activity analysis with the aim to differentiate between the activity recorded during depression, mania and euthymia.

Chapter 2

Bipolar disorder

Bipolar disorder, formerly known as manic depression, is a mental illness that causes unusual shifts in mood and energy. The affected person experiences prolonged periods of felling high and full of energy, called mania, as well as episodes of mania's exact opposite, depression. It is a chronic disorder with episodic progress. It has been estimated that bipolar disorder has a lifetime prevalence of 2.4% in population worldwide [1]. The same study also reported high comorbidity of bipolar disorder with other mental disorders, particularly anxiety disorders, behaviour disorders, and substance use disorders.

Cognitive impairment has been reported among those affected by bipolar disorder [2]. Furthermore, it was reported [3] that individuals with bipolar disorder experience 41.2 days out of role per year on average, that is, days when a person cannot carry on with their work or perform normal activities due to the complications caused by a disorder. Patients with bipolar disorder have one of the highest risks of death by suicide, which was estimated to be 20–30 times higher than in the general population [4].

2.1 Episodes

Bipolar disorder and its episodes are usually diagnosed in accordance with the Diagnostic and statistical manual of mental disorders (DSM-V) [5] that describes the main characteristics of each of the episodes. These are presented in the following subsections.

2.1.1 Major depressive episode

A depressive episode is characterized by strong and persistent feelings of hopelessness and sadness. The person undergoing this episode might experience disinterest in activities he or she would otherwise find pleasurable. A significant drop in energy and physical activity is also one of the symptoms attributed to depression and an individual is thought to be more prone to display periods of inactivity during the daytime. The period during which the symptoms of depression are present must last at least two weeks in order to be recognized as a major depressive episode.

Despite the fact that the symptoms of a depressive episode in bipolar disorder are very much like those of a depressive episode in unipolar depression, there are some subtle differences between the two. Depressive episodes in bipolar disorder are more frequent and have a shorter duration than those in unipolar depression. Moreover, a person undergoing a bipolar depressive episode is more likely to experience hypersomnia and psychosis [6].

2.1.2 Manic episode

The state of mania is associated with elevated mood, decreased need to sleep and high irritability. There is also a visible increase in energy levels and physical activity. A person can feel like he or she can handle doing multiple things at the same time, while not being able to properly focus on

any of them. Patients might also experience psychotic sympyoms [7]. The symptoms of mania must last at least one week in order for them to be recognized as a manic episode.

Patients usually show poor judgment and make irresponsible decisions that are not in their nature and that they would not make otherwise. These carefree actions can, for example, endanger their personal relationships, put them in a complicated financial situation or even put their health at risk.

2.1.3 Hypomanic episode

A hypomanic episode is very similar to a manic episode, but the change in mood and behaviour is not as clear as it is in mania. The patient is less likely to notice the change in his behaviour, but the change is still noticeable by other people. The symptoms of hypomania must last at least 4 days in order to be recognizd as a hypomanic episode. Differentiating between mania and hypomania results in recognizing different types of bipolar disorder, as described in the following section 2.2.

2.2 Classifications

Patients with bipolar disorder are further classified into separate types of bipolar disorder. These are distinguished by the types of episodes suffered by a patient. The criteria for the diagnosis of these types are described in DSM-V [5].

2.2.1 Bipolar I disorder

The bipolar I disorder diagnosis can be made if at least one manic episode has occured during the lifetime of a patient. It is not unusual for other episode types to occur as well, although they are not required in order to perform the diagnosis.

2.2.2 Bipolar II disorder

The difference between bipolar I and bipolar II disorders is that the diagnosis of bipolar II disorder does not require an occurence of a manic episode. However, it is required for a patient to undergo both a major depressive episode as well as an episode of hypomania. The frequency of episode occurence is higher in bipolar II disorder than in bipolar I disorder [8].

2.2.3 Cyclothymic disorder

The cyclothymic disorder involves occurence of multiple periods during which the symptoms of hypomania are present and periods during which the patient shows the symptoms of a major depressive episode. However, these symptoms fail to fully meet the criteria for the periods to be diagnosed as either of the episodes. These periods must persistently occur for at least 2 years.

2.3 Diagnosis

It is important to classify bipolar patients and to assess their mood state correctly, as the treatment depends on wheter the patient is undergoing an episode or is in a remission. Furthermore, the treatment approach differs between the individual episode types [7]. The clinical diagnosis made by a psychologist remains vital, as there is currently no known valid biomarker for this condition, although there are some promising propositions [9].

An interview with the patient is conducted and the history of the patient's mood state must be investigated in order to assess past symptoms of bipolar disorder and mood swings. An

interview with the patient's close relatives might also prove to be beneficial while performing the patient's diagnosis [7].

The severity of a manic episode can be assessed using the Young Mania Rating Scale (YMRS) [10] and the severity of a depressive episode can be assesed using either the Montgomery-Asberg Depression Rating Scale (MADRS) [11] or the Hamilton Depression Rating Scale (HDRS) [12].

It is challenging to tell bipolar disorder apart from unipolar depression, as the depressive episode often occurs at the onset [13] of the illness. Consequently, many patients are misdiagnosed until the first occurence of mania in bipolar I disorder or hypomania in bipolar II disorder, which is harder to recognize due to less pronounced symptoms.

2.4 Treatment

2.4.1 Acute treatment

The main aim while treating an acute episode of bipolar depression is to make sure that the patient and the people in his surroundings are safe and to stabilize the patient's mood and reestablish the state of euthymia.

The acute treatment is very difficult, since it is possible that some treatments of mania will induce a depressive episode and treatments of depression will induce mania.

The acute episodes of both mania and depression are mainly treated using antipsychotics and mood stabilizers. Antipsychotics were found to be more effective in treating mania than mood stabilizers [14] but it is also possible to combine these two medications [7]. Antidepressants are sometimes used for treating bipolar depression, however, their usage is very controversial, since it might induce a switch to mania [15].

2.4.2 Long-term management

The goal in the long-term management of bipolar disorder is to prevent a relapse. This is achieved by a combination of medication and therapy. Lithium is often prescribed as a mood stabilizer as it helps to prevent the occurence of both manic and depressive episodes [7]. In addition to medication, it is also favourable to undergo a psychotherapy treatment, for example family-focused therapy, cognitive-behavioural therapy or psychoeducation [14].

Chapter 3

Actigraphy

As has been already stated in the chapter about bipolar disorder, a change in physical activity and energy is an important symptom that occurs during bipolar disorder episodes and is also one of the criteria for their diagnosis. Therefore, movement recordings might provide us with insightful information that could be useful for the diagnosis of the patient's state and help us assess the course of the illness.

One of the techniques that are most commonly used to record physical activity is called actigraphy. Actigraphy uses an accelerometer-based device called an actigraph in order to record movement. This device is usually worn on the wrist, but can be possibly also worn on the ankle or around the waist. It is a non-invasive method that does not incur any major restrictions on the actigraph wearer, and as such has the ability to record activity patterns that are most true to the patient's regular activity. Figure 3.1 below contains an example of an activity record obtained using actigraphy.

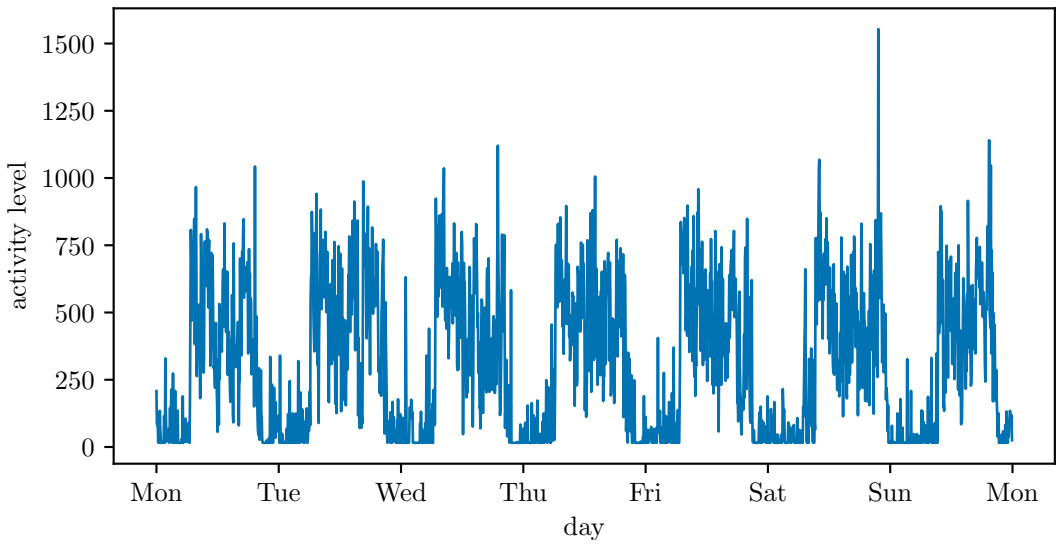

Figure 3.1: An excerpt of patient 808's recordings, downsampled to have a sampling period of 5 minutes in order to improve clarity.

The first documented use of an actigraphic device dates back to 1959 when a modified Omega self-winding wristwatch, referred to as an actometer, was used in an attempt to objectively evaluate the activity in hyperactive children [16].

Although the use of actometer sounds very interesting and has been quite innovative, the device was later proven to be rather unreliable [17].

Fortunately, actigraphs, along with electronic devices, have since then undergone considerable development, and the activity measurement process has been digitized. Modern actigraphic devices are precise, offer high sampling frequencies, and have enough memory to store records in order of months, depending on the selected sampling frequency. They present a useful diagnostic tool.

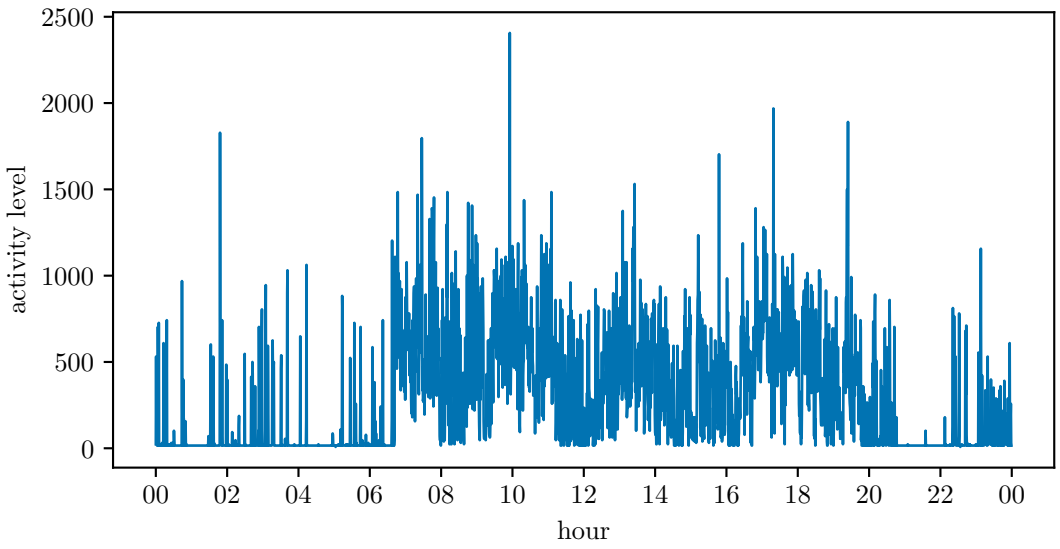

Figure 3.2: Data recorded on Monday from the identical extract as in 3.1, displayed at the original sampling period of 30 seconds.

Actigraphy have been mainly employed in order to assess the sleep-wake patterns in people with sleep disorders, where it has been preferred over polysomnography, especially due to the possibility to conveniently record data over longer periods of time [18].

Actigraphs have also been used to study circadian rhythms and motor activity in affective disorders, including bipolar disorder [19], [20].

Over the years, multiple methods for the analysis of actigraphy data have been used and have proven to be able to extract information of significant value. These methods range from simple activity mean calculations to cosinor analysis and nonparametric analysis, which will be presented in chapter 6.

Chapter 4

Available data

The data analyzed in this bachelor's thesis consisted of longitudinal monitorings of bipolar disorder patients coupled with mood surveys. The durations for which these patients were monitored varied, with the longest record of nearly 3 years and the shortest having a length of 6 months. The median of the recording lengths was 2 years. The data was provided by my supervisor and were collected as part of an ongoing research organized by the National Institute of Mental Health in Klecany.

Patients included in this study wore an actigraphic device that countinuously recorded their activity. The sampling period at which the data was processed was 30 seconds. In addition to wearing an actigraph recording their motor activity, patients filled out questionnaires designed to assess their mood state. The questionnaires were submitted at irregular intervals (approximately once every 1–2 weeks) using a mobile application and patients were instructed to fill them out with respect to their mood over the course of the previous week. The questionnaire was composed of multiple statements regarding the subject's mood state and overall well-being.

There were four statements focused on detecting the symptoms of a depressive episode and other four statements that were meant to ascertain the main symptoms of mania. There were also two less specific statements assessing an overall change in the patient's state.

The participants rated each of the statements in order to express their level of agreement with the statement. The possible rating values ranged from 0, as in "I do not agree", to 4, as in "I completely agree". After its submission, each questionnaire was assigned a level of severity of both the depressive and the manic symptoms. Level 0 indicated none or mild symptoms, level 1 indicated a moderate level of symptoms, and level 2 was assigned to those questionnaires that displayed severe episode symptoms. It can be expected that the data obtained through this survey is far more subjective than the activity data collected using actigraphy.

A significant difference in the number of depressive and manic episodes was found based on the submitted questionnaires, with depression being collectively the more prevalent episode type. This finding is in accordance with the fact that even though episodes of mania and depression have similar occurence rate, depressive episodes generally have a longer duration [21]. As a result, it is far more probable that more questionnaire submissions indicating the presence of depressive episode symptoms will be collected. Nonetheless, some of the patients displayed a more balanced count of episode types with no visible predominance of either of the two episode types.

As it is with most real-world data collection surveys, the obtained datasets contained missing values. These can be attributed to a number of possible reasons, for example noncompliance of the patient, device maintenance or health-related complications. The exact reasons were not investigated.

Furthermore, data segment labels indicating the type and length of an episode that a patient was undergoing during a certain point in time were provided. These labels were assigned by hand in cooperation with the psychologists who used the YMRS and MADRS scales to asssess

the severity of the patient's symptoms.

In conclusion, two sources of data were available: the objective recordings of physical activity obtained through the use of actigraphy with a sampling period of 30 seconds, and the more subjective self-reports of the patient's current mood.

Chapter 5

Artificial data generator

I designed a generator of artificial actigraphic data in order to perform experiments and study the behaviour of the methods that were used for the data analysis. The artificial data is constructed using a square wave signal, simulating the circadian activity rhythm, with addition of white noise, representing the changing motor activity.

The generator provides a number of parameters that can be used to create a dataset with the required properties. These parameters include duration and level of activity during the day, percentage of time spent resting during the day, nocturnal disturbance probability and intensity, and sleep duration. The mean and standard deviation is set for each of these variables and the concrete values used when generating a single day-night period are sampled from a normal distribution.

Based on the set parameters, a chosen number of day-night periods is generated. A tool that allows to conveniently create datasets comprised of periods that were generated using different settings was developed as well.

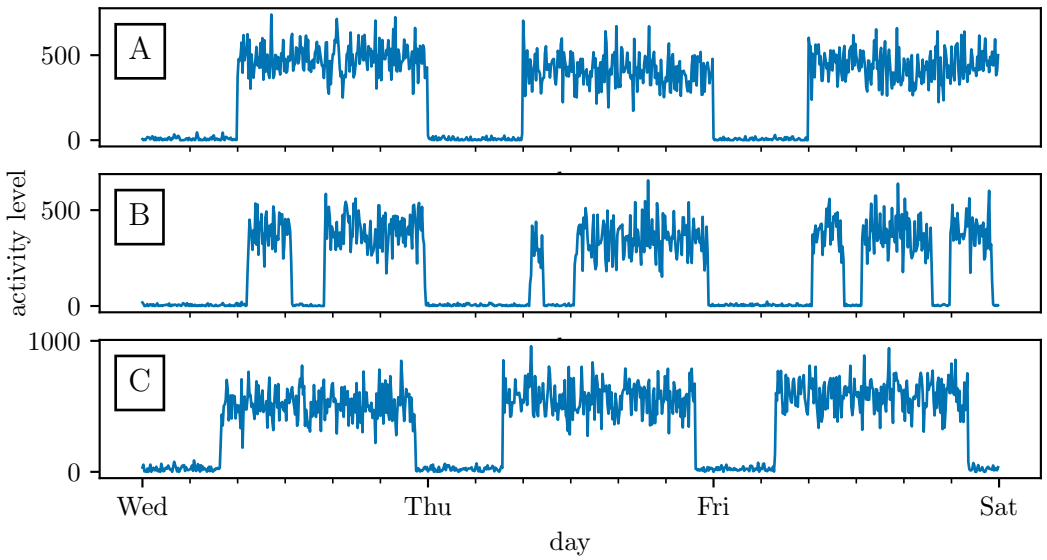

Figure 5.1: Examples of generated datasets. The recordings in figures (A), (B), and (C) are meant to represent activity during euthymia, depression, and mania, respectively.

I prepared six different generator settings that were meant to approximately represent the expected activity during each of the bipolar episodes and during an euthymic state. Certain parameters were hand-picked so that activity levels during the day and the level of night dis-

turbances reflected the real-world collected data. Two different settings were created for each of the mood states.

The euthymic settings were set to be fully synchronized with the 24-hour cycle and to experience a low level of night disturbances. No drop in activity was set to occur during the day. The settings representing a depressive episode were selected so that a more fragmented rest-activity rhythm was apparent, with periods of rest taking place during the daytime. They were also set to have varied durations of activity and sleep and to have a lower mean of activity level than the euthymic settings. The manic activity generator settings had the highest activity level mean and the highest probability of nocturnal disturbances. They were also set to have a shorter and more varied sleep duration than the other settings. One of the manic settings was configured so that the resulting data would be poorly synchronized with the 24-hour cycle.

Chapter 6

Nonparametric actigraphy analysis

After obtaining the activity recording, a question presents itself. How do we process it in order to extract features that will help us interpret the captured data? Various methods have been used with this question in mind, most notably cosinor analysis and nonparametric analysis.

Using these methods, it is possible to study the circadian activity rhythm in the observed subject. It has been found that both the both rest-activity rhythm and sleep-wake rhythm are disrupted in bipolar disorder as well as in other mood disorders [19], [22].

Nonparametric analysis utilizes several variables to study the circadian activity rhythm, namely intradaily variability, interdaily stability, M10, L5, and relative amplitude. This approach was proposed in an article by Witting et al. [23] that studied the alteration in the circadian activity rhythm in aging and Alzheimer's disease and has been quite popular since then.

This method of data analysis is called nonparametric, because the variables it uses to study the rest-activity rhythm are not associated with parameters of a known function [24]. Therefore no assumptions are made about the actigraphy signal as to its form or distribution.

6.1 Intradaily variability

The intradaily variability (IV) is a measure of the rest-activity rhythm fragmentation. Higher values of IV indicate more frequent and rapid transitions between periods of rest and activity. Therefore the presence of daytime nappings or nocturnal activity in the recording leads to an increase in IV.

Significant differences in IV have been found for example between populations of different age, with IV achieving higher values in the elderly than in their younger counterparts [25]. Higher values of IV were also found in patients with Alzheimer's disease compared to controls of similar age [23].

The intradaily variability is computed as

$$\mathrm{IV} = \frac{\text{mean square successive difference}}{\text{variance}} = \frac{N \sum_{i=2}^{N} (X_i - X_{i-1})^2}{(N-1) \sum_{i=1}^{N} (X_i - \bar{X})^2} \tag{6.1}$$

where X_i is the activity level recorded at time i, $\bar{X}$ is the mean activity level, and N is the total number of recorded points.

Figure 6.1 showcases the difference in IV between two activity recordings. Recording (A) achieves smaller IV value than recording (B). Notice that the activity in recording (B) is much more fragmented than that in recording (A), with periods of rest occuring throught the day. For example, there is a visible pattern of a short activity early in the morning followed by a period of rest, and spans of inactivity in the afternoon. On the other hand, the activity in recording (A) is much less variable and does not show any significant drops in activity throughout the

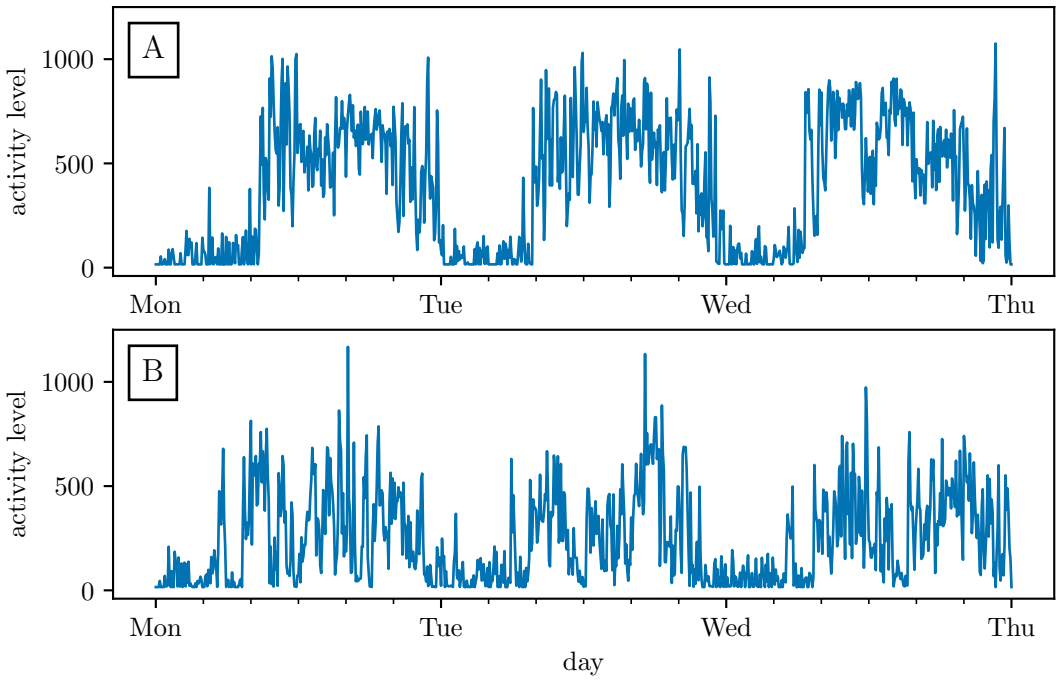

Figure 6.1: IV sensitivity to transitions between rest and activity. Recording displayed in figure (A) achieves a smaller IV value, indicating smaller activity fragmentation, than the recording displayed in figure (B).

daytime.

Intradaily variability was usually computed using hourly sampled data, but as the technology of actigraphy devices improved and made it feasible to record activity at higher sampling rates, so has the approach to computing this variable changed. It has been proposed that computing the IV for data resampled for each of minute samplings in range from 1 minute to 60 minutes and taking the mean of these values could a more sensitive measurement called IVm—intradaily variability mean [24]. This variable was therefore computed and noted as well, along with IV values of data resampled to individual samplings from the mentioned range. These IV values will be reffered to by appending the sampling period of data from which they were computed. For example IV60 will refer to IV of data that was resampled to a sampling period of 60 minutes.

6.2 Interdaily stability

The interdaily stability (IS) signifies the subject's circadian rhythm stability and how well he or she is synchronized to the 24 hour cycle, with higher values indicating better synchronization. Deviations in either activity or sleep onset and duration are responsible for lower values of interdaily stability.

Lower interdaily stability values were reported in bipolar disorder patients than in healthy controls [19] and IS has also been used to study the effect of light exposure on the circadian rhythm in Alzheimer's disease [26].

The interdaily stability is computed as

$$\mathrm{IS} = \frac{\text{average daily profile variance}}{\text{variance}} = \frac{N \sum_{h=1}^{p} (X_h - \bar{X})^2}{p \sum_{i=1}^{N} (X_i - \bar{X})^2} \tag{6.2}$$

where X_h is the activity level corresponding to the point of the average daily profile at time h, p is the number of points recorded per day, and X_i, $\bar{X}$, and N refer to the same variables as in the intradaily variability equation 6.1.

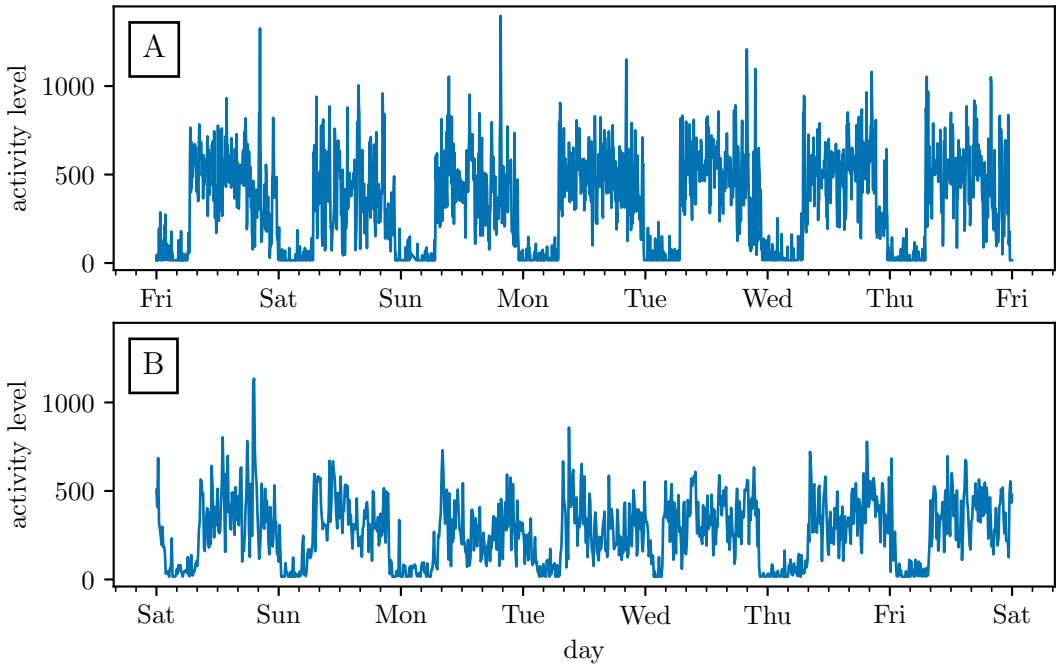

Figure 6.2: IS sensitivity to transitions between rest and activity. Recording displayed in figure (A) achieves a higher IS value, indicating better synchronization with the 24-hour day cycle than the recording displayed in figure (B).

Figure 6.2 gives an example of two different recordings. Recording (A) has a visibly stable rhythm with no significant shifts in either sleep or activity onsets. On the contrary, recording (B) exhibits a less stable circadian rhythm with apparent changes in sleep onsets and durations, with the most notable deviation occuring during the night from Tuesday to Wednesday. The IS values computed for both of these recordings reflect these notions, with recording (A) achieving a higher value than recording (B), indicating a more stable circadian rhythm.

Interdaily stability mean—ISm—has been computed as well, but the range of samplings considered for this computation differed from that used while computing IVm. Included were only those minute samplings from the range of 1 minute to 60 minutes that evenly divide the number of minutes in a full day. This restriction was imposed so that each full day included in the activity recording contributes equal number of data points during the evaluation of the average daily profile and therefore that no bias is introduced.

6.3 M10, L5, and relative amplitude

Three more variables are considered as well along with IV and IS. First of them is M10, which quantifies the mean activity during the 10 most active consecutive hours in an average daily profile. Similarly, L5 is the mean activity during the 5 least active consecutive hours of the average daily profile. The main purpose of the variables M10 and L5 is to capture the activity level during the day and during the night, respectively. The times at which the most and the

least active periods began were taken note of as well.

These two new variables can be further combined in order to compute the relative amplitude (RA), which was not studied in [23].

$$\mathrm{RA} = \frac{\mathrm{M10} - \mathrm{L5}}{\mathrm{M10} + \mathrm{L5}} \tag{6.3}$$

From equation 6.3, it is apparent that a higher difference between M10 and L5 leads to a higher value of RA.

6.4 Effect of missing data

As mentioned in chapter 4, the analyzed data contains missing values. This fact makes the data analysis more complicated, as incomplete data can cause serious errors in the values of estimated statistics, including the nonparametric analysis variables. This problem could be solved by simply disregarding the records containing any missing values. However, this would significantly scale down the volume of available data, and therefore was not considered an option. For this reason, I performed an experiment that would help assess the effect of missing data on the nonparametric variables using complete available records and the artifical actigraphy data generator.

6.4.1 Simulation of missing data

I took advantage of both the available data and the artificial data generator in order to simulate the missing values in actigraphy recordings and to estimate the severity of error they introduce to the values of nonparametric variables. Both types of data, real and artificial, was processed individually and then compared.

I selected two week-long records from the data collected using actigraphy that did not lack any data points for each of depression, mania and euthymia. Consequently, each of these states was equally represented during this experiment. Self reported mood was used to estimate the state of the patient during the extracted recording.

The artificially generated data was acquired using the six different generator settings, where once again each of the mood states corresponded with two of the used settings. Data was generated during the experiment and had not been produced beforehand.

Missing segments of data were simulated in the following fashion. A record was randomly selected and its values of nonparametric variables were computed. After that, the record was divided into a chosen number of disjoint parts of equal size and random continuous data segments of equal size were removed from each of the parts. The size of the removed data segments depended on the total required percentage of the removed data.

The values of nonparametric variables were then computed for the edited record and the relative error in the resulting value was computed and expressed in percentage of the value computed for the full record.

In the case of artificial data, a random generator setting was selected out of the six available options and used to generate a week-long recording. Missing values were simulated using the same process as in the case of real data and the error was again measured using the relative error.

These steps were repeated 1,000 times for each of the considered total percentages of missing data, which ranged from 5% to 50%. The chosen numbers of removed segments were 7, resulting in the removal of longer data segments from every day of the record, and 14, which produced shorter missing data segments occuring each half-day.

6.4.2 Experiment results

All the experiment results are displayed as relative error means along with their 95% confidence intervals.

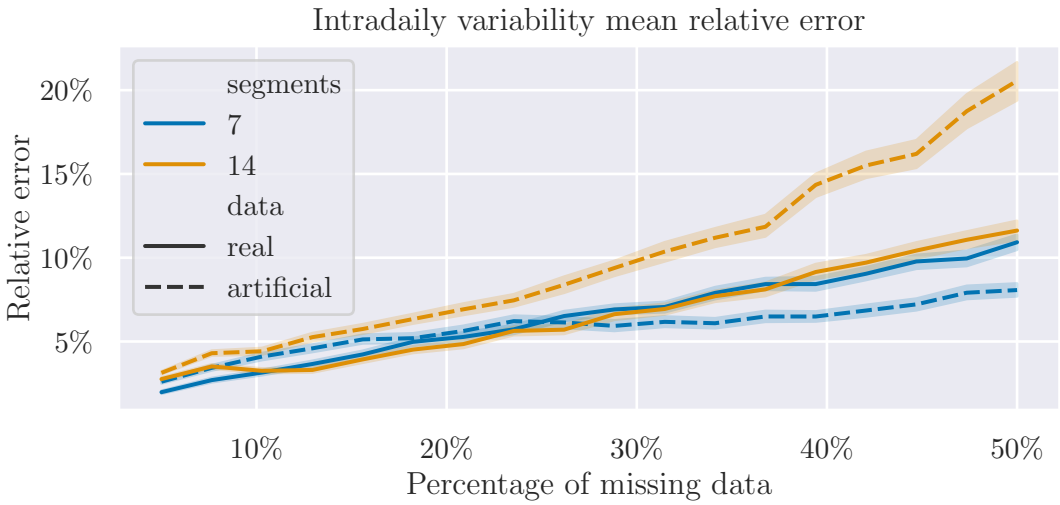

Figure 6.3: Susceptibility of IVm to missing data.

Figure 6.3 displays the results of the experiment for the variable IVm. It can be seen that the negative effect of missing values in real data seems somewhat linear and is similar for both chosen numbers of removed data segments across the whole range of simulated percentages of absent data points. The relative error attains a value of 5% at 20% of missing data, and steadily climbs up to a value of 11.5% at 50% of missing data. On the other hand, the errors computed using artificial data diverge quickly at around 25% of missing data. The relative error grows more quickly in the data where smaller segments were missing more frequently and attains a significantly higher value than in the data missing longer segments more sparsely.

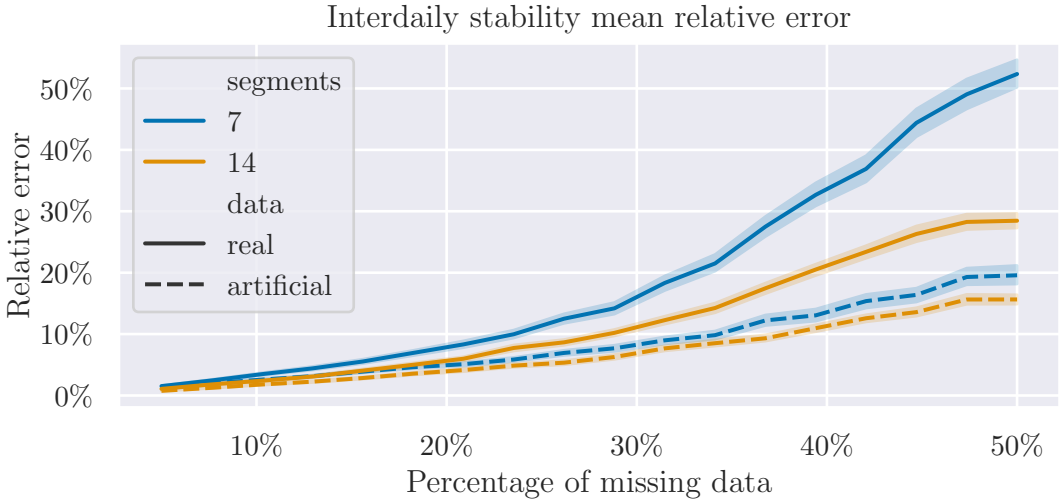

Figure 6.4: Susceptibility of ISm to missing data.

The relative error in the values of ISm can be seen in figure 6.4. All the configurations behave similarly up to 20% of missing data. After that, the error in real data was higher when 7 longer sections of data were missing than when 14 sections were not available. The artificial

data confirms that the error is higher when longer periods of data are missing. On the other hand, it undervalues the level that both errors can attain.

The error in RA in artificially generated data mimics that in real data and is virtually indistinguishable up to 30% of missing data, but ultimately plays down its value. However, this might very well be due to the distribution of missing data segments. The fact that a portion of data is removed in each half-day of the recording when the number of removed segments is 14 means that there is a higher probability that both the L5 and M10 variables will be affected by the missing values. Consequently, their errors will compound during the computation of RA.

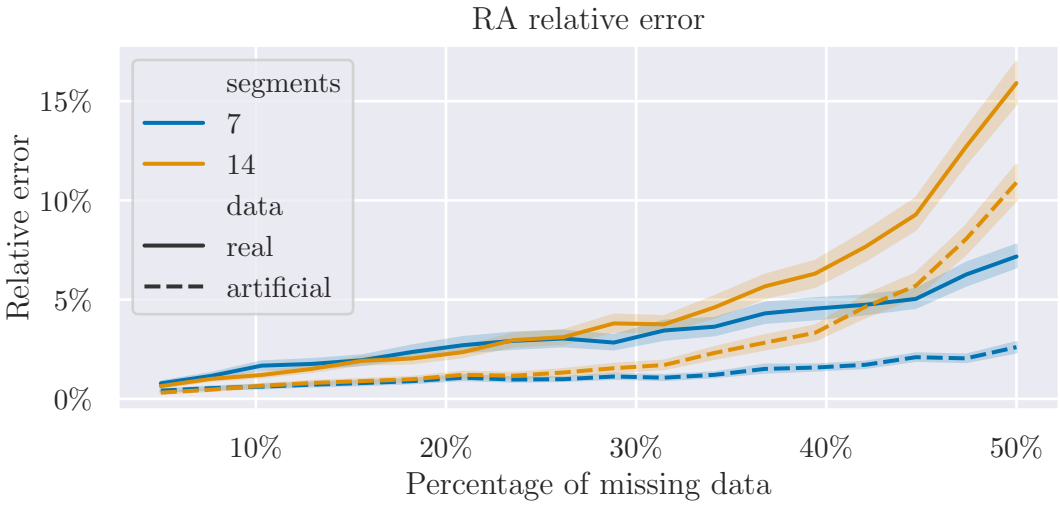

Figure 6.5: Susceptibility of RA to missing data.

This simple inspection provides us with at least some notion of what we can expect when employing the tools of nonparametric analysis together with actigraphy recordings that lack certain amount of data points. According to the results of this experiment, the ISm variable is more error-prone than IVm after the introduction of unavailable data measurements into the datasets.

6.5 Comparison with cosinor analysis

Cosinor analysis is another popular technique for studying circadian rhythms. It is based on fitting a parametrized cosine curve, shown in equation 6.4, to observed data using the least squares method [27].

$$Y = M + A\cos(\frac{2\pi t}{\tau} + \phi) \tag{6.4}$$

Based on the fit of this curve, four characteristic parameters of the circadian rhythm can be obtained, MESOR M, amplitude A, period τ, and acrophase ϕ. MESOR M is a measure of the rhythm-adjusted mean activity. Amplitude A shows the difference between the MESOR and the maximum achieved level of activity during a cycle, parameter τ specifies the duration of one cycle, and acrophase ϕ is the time at which the amplitude of the fitted curve occurs.

Cosinor analysis can be a very useful tool when the analyzed data does not consist of equidistant measurements of activity. However, due to the assumption that the circadian activity rhythm follows a cosine curve, it does not capture certain features of physical activity, for example the rhythm fragmentation.

It has also been shown, that nonparametric analysis variable IS is more sensitive and precise when inspecting the circadian rhythm stability than any other of the cosinor analysis variables

[26].

Even though cosinor and nonparametric analyses are the still the most common methods used to analyze actigraphy data, novel approaches are still emerging [28].

Chapter 7

Sample entropy

The concept of entropy with relation to information was introduced by Claude Shannon in his famous article "A Mathematical Theory of Communication" [29], which resulted in the creation of a new mathematical field of information theory. Information entropy can be thought of as an amount of uncertainty associated with a random variable.

However, computing the entropy of real-world time series datasets has been quite problematic, due to their finite length and a frequent presence of noise [30]. That was until Pincus et al. introduced [30] a new statistic called approximate entropy which was developed in order to address these specific problems.

Approximate entropy of time series quantifies the amount of regularity or complexity in the observed data [30]. Time series containing higher count of similar patterns are recognized as having higher regularity and therefore achieve lower values of entropy.

Sample entropy, developed by Richman and Moorman [31], emerges as a modification of approximate entropy. These modifications were made in order to address certain problematic behaviours of approximate entropy, most importantly the bias towards lower values of entropy introduced during the process of its computation.

Assessing the regularity can be of high interest in the analysis of biomedical signals. These entropy measures were used for example to analyze ECG data in Alzheimer's disease patients [32] and neonatal heart rate variability [33].

Based on the description and characteristics of episode types in bipolar disorder, it can be expected that the complexity of motor activity will differ between these types and euthymia. In comparison with entropy values observed during euthymia, the entropy values could attain higher levels during mania due to its impulsive and chaotic nature and lower levels during episodes of depression due to more frequent periods of inactivity.

7.1 Computation

Given that these two statistics, approximate and sample entropy, are closely related, the computation of approximate entropy as defined in [30] will be explained and then in turn the computation of sample entropy with accordance to [31] will be presented. The main differences between these two measures will be highlighted.

7.1.1 Approximate entropy computation

Given a time series $\mathbf{u} = (u_1, \ldots, u_N)$ consisting of N data points, and two parameters, the embedding dimension m, and the tolerance r, we compute approximate entropy as follows.

Let $\mathbf{x}_m(i) = (u_i, \ldots, u_{i+m-1})$ denote the vector (sometimes referred to as a template) of m consecutive data points of the time series commencing with the data point at the index i. Construct all such templates for every i from 1 through $N - m + 1$ and define the distance

$d[\mathbf{x}_m(i), \mathbf{x}_m(j)]$ between these vectors as the Chebyshev distance $d(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|$, the maximum difference of respective scalar components of the vectors.

Afterwards, for every template $\mathbf{x}_m(i), i = 1, \ldots, N + m - 1$ compute the value

$$C_i^m(r) = \frac{|\{\mathbf{x}_m(j) \mid d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r\}|}{N - m + 1} \tag{7.1}$$

where $j = 1, \ldots, N - m + 1$. Every vector $\mathbf{x}_m(j)$ that is within r of $\mathbf{x}_m(i)$, that is for which it holds that $d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r$, is called a template match with $\mathbf{x}_m(i)$. The value $C_i^m(r)$ can be interpreted as the probability of occurence of a pattern that is similar to the template $\mathbf{x}_m(i)$ up to a tolerance r.

Having computed these values, define

$$\Phi^m(r) = \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N - m + 1} \tag{7.2}$$

Subsequently, the approximate entropy parameter is defined as

$$ApEn(m, r) = \lim_{N \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \tag{7.3}$$

Given a finite number of data points, we can estimate the approximate entropy as

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{7.4}$$

By inputting the definition of the variable $\Phi^m(r)$ into the previous equation 7.4 and performing a few simple algebraic manipulations, we can see that the approximate entropy can be estimated as

$$\begin{aligned} ApEn(m, r, N) &= \Phi^m(r) - \Phi^{m+1}(r) \\ &= \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N - m + 1} - \frac{\sum_{i=1}^{N-m} \ln C_i^{m+1}(r)}{N - m} \\ &\approx \frac{\sum_{i=1}^{N-m} \left[\ln C_i^m(r) - \ln C_i^{m+1}(r)\right]}{N - m} \\ &= \frac{1}{N - m} \sum_{i=1}^{N-m} \ln \frac{C_i^m(r)}{C_i^{m+1}(r)} = -\frac{1}{N - m} \sum_{i=1}^{N-m} \ln \frac{C_i^{m+1}(r)}{C_i^m(r)} \end{aligned} \tag{7.5}$$

Expressing the approximate entropy using this equation allows us to interpret it as the negative average of the logarithms of conditional probabilities, that given that a vector sequence is similar with a template $\mathbf{x}_m(i)$ for m consecutive data points, it remains similar for an additional successive data point.

Problems could arise during this computation, in case that no match for a template $\mathbf{x}_m(i)$ was found. In that case, the value of $C_i^m(r)$ would equal zero, and an undefined $\ln 0$ would occur while computing $\Phi^m(r)$ in the equation 7.2. However, these complications are mitigated by considering every template $\mathbf{x}_m(i)$ to trivially match itself, and therefore at least one match is found for every template and the values of $C_i^m(r)$ are sure to be greater than zero. This is also true for the templates $\mathbf{x}_{m+1}(i)$. Unfortunately, this precaution introduces an unwanted bias of the approximate entropy towards estimating greater regularity than is truly present, indicating a less complex underlying system.

The sample entropy reduces this bias by excluding the self matches and by taking a different approach to estimating the conditional probabilities identified in the equation 7.5.

7.1.2 Sample entropy computation

The sample entropy takes, much like the approximate entropy, two parameters, the embedding length m and the tolerance r. Given a data series $\mathbf{u} = (u_1, \ldots, u_N)$ of N data points, embedding dimension m and tolerance r, we can compute the sample entropy as follows.

As in the computation of approximate entropy, once again define the vector of m consecutive data points starting with the point at index i $\mathbf{x}_m(i) = (u_i, u_{i+1}, \ldots, u_{i+m-1})$, and define

$$d\left[\mathbf{x}_m(i), \mathbf{x}_m(j)\right] = \max_{k=0,\ldots,m-1} |u_{i+k} - u_{j+k}|$$

.

Compute the number $B_i^m(r)$ for every template $\mathbf{x}_m(i)$ in range $i = 1, \ldots, N-m$, which equals the number of template matches for the specific template, excluding the self-match, divided by the total number of templates of length m.

$$B_i^m(r) = \frac{|\{\mathbf{x}_m(j) \mid d\left[\mathbf{x}_m(i), \mathbf{x}_m(j)\right] \leq r, i \neq j\}|}{N - m + 1} \tag{7.6}$$

Essentially, the number $B_i^m(r)$ represents the probability that any other vector will match with the template $\mathbf{x}_m(i)$. This variable is analogic with the variable $C_i^m(r)$ defined in 7.1, but there are some important differences.

Firstly, it is not allowed for the vector to match with itself and by doing so the introduction of a bias towards lower complexity is prevented. Secondly, notice that only the first $N - m$ templates of length m are considered when estimating the probability. The reason for this is that even though there are $N - m + 1$ templates of length m, there are only $N - m$ templates of length $m + 1$, and therefore the last template of the length m does not have any parallel template of length $m + 1$ that would start at the same index.

Having computed the variables $B_i^m(r)$ for $i = 1, \ldots, N - m$, define $B^m(r)$ as their average.

$$B^m(r) = \frac{\sum_{i=1}^{N-m} B_i^m(r)}{N - m} \tag{7.7}$$

The resulting value is the probability that any two sequences will match for m consecutive data points. It is the counterpart of $\Phi^m(r)$ in the equation 7.2, with the distinction that the natural logarithm is not applied here, as it is when computing $\Phi^m(r)$.

Similarly define $A_i^m(r)$ as the number of template matches $\mathbf{x}_{m+1}(j)$ for the specific template $\mathbf{x}_{m+1}(i)$, excluding the self-match, divided by the total number of templates of length m. Notice, that instead of comparing the vectors of length m, we are now comparing the vectors of length $m + 1$, thus we include one additional data point in the comparison.

$$A_i^m(r) = \frac{|\{\mathbf{x}_{m+1}(j) \mid d[\mathbf{x}_{m+1}(i), \mathbf{x}_{m+1}(j)] \leq r, i \neq j\}|}{N - m + 1} \tag{7.8}$$

As in the previous case, the probability that any two sequences match for $m + 1$ points, $A^m(r)$, is computed.

$$A^m(r) = \frac{\sum_{i=1}^{N-m} A_i^m(r)}{N - m} \tag{7.9}$$

The sample entropy parameter is then defined as

$$SampEn(m, r) = -\lim_{N\to\infty} \ln \frac{A^m(r)}{B^m(r)} \tag{7.10}$$

which for a finite length time series data is estimated as

$$SampEn(N, m, r) = -\ln \frac{A^m(r)}{B^m(r)} \tag{7.11}$$

This equation can be further simplified in order to provide a more insighfult interpretation of the sample entropy. Assign B to equal the total number of template matches of length m, and A to equal the total number of template matches of length $m+1$.

$$\begin{aligned} B &= \frac{(N-m+1)(N-m)}{2} B^m(r) \\ A &= \frac{(N-m+1)(N-m)}{2} A^m(r) \end{aligned} \tag{7.12}$$

Using these new variables, we can see that

$$SampEn(N,m,r) = -\ln \frac{A^m(r)}{B^m(r)} = -\ln \frac{A}{B} \tag{7.13}$$

Sample entropy is therefore the negative natural logarithm of the conditional probability that given that two vector sequences are similar for m consecutive data points, they also remain similar for the next data point. Sample entropy estimates this conditional probability for the whole time series, whereas the approximate entropy does so for each of the templates $\mathbf{x}_m(i)$ individually.

Because the template self matches are excluded from the computation of sample entropy, it is possible that the resulting value will not be defined due to the occurence of $\ln 0$ in 7.11. However, this situation would arise only if no matches would have been found either for any template of length m, resulting in $B^m(r) = 0$, or for any template of length $m+1$, resulting in $A^m(r) = 0$.

7.2 Usage and behaviour

Now that it has been shown how the sample entropy is computed, it is possible to take a closer look at how the parameters m and r affect its behaviour and how these parameters should be chosen. I will also provide an example showcasing the ability of sample entropy to differentiate between signals of varying complexity.

7.2.1 Choosing the parameters

Pincus et al. reported [30] good results of approximate entropy when used with settings where $m = 2, 3$ and r was set to be in the range of 0.1 to 0.25 of the standard deviation of the processed data. This allowed for the approximate entropy to find enough template matches so that the conditional probabilities could be reliably estimated, while also preserving its discriminative ability. These recommendations have also been transfered to the usage of sample entropy and many researchers and practitioners abide by them to this day.

Although these settings might work well in many cases, different settings should be considered as well, preferably on a case by case basis. There are no strict guidelines for choosing the parameters. It might prove useful to perform an inspection of the possible values for r and m as a more optimal setting could be found when the nature of the analyzed data is taken into consideration.

Note that every time a value of the parameter r is displayed or mentioned throughout the remainder of this thesis, it specifies the multiplicator of the standard deviation of the analyzed data used to obtain the tolerance r utilized during the entropy computation, unless stated otherwise.

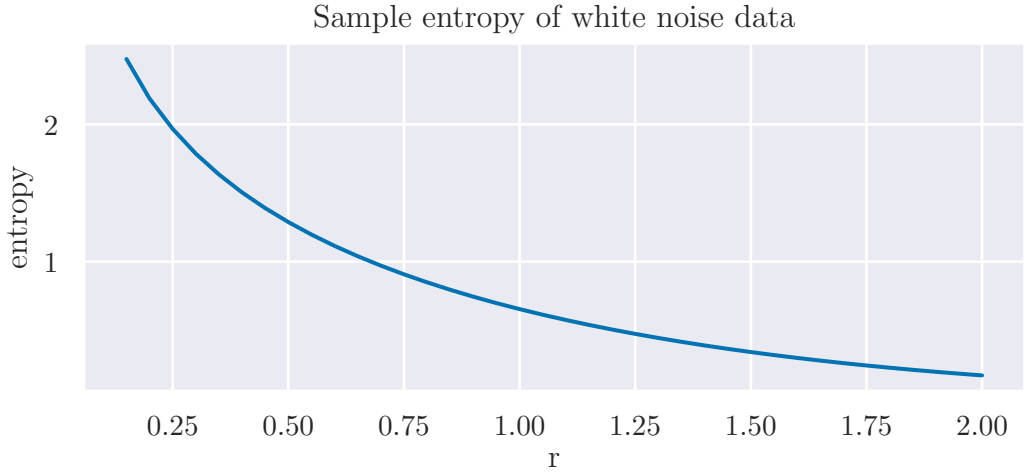

Figure 7.1: Behaviour of the sample entropy with varying tolerance r. Computed using white noise generated by sampling 10,000 data points from the normal distribution.

7.2.2 Behaviour with regard to the parameters

Figure 7.1 shows the behaviour of sample entropy with varying parameter r and fixed $m = 2$. Tolerance r can be thought of as a noise filter. That is why it is frequently chosen to equal a multiple of the signal's standard deviation, as this allows it to disregard small differences between vector templates and still consider them to be similar. As r grows, the requirement $d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r$ becomes less strict and it is easier for the templates to match. The increase in matches for both templates of length m and $m + 1$ results in lower entropy values.

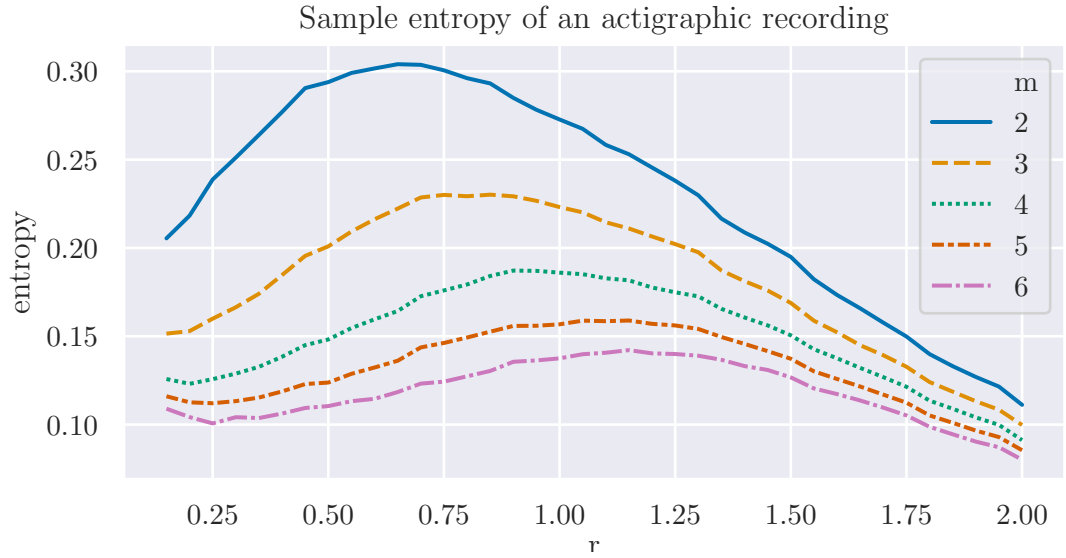

Figure 7.2: Behaviour of sample entropy with varying tolerance r and embedding dimension m. Computed using 7 days of actigraphy recordings with sampling period of 30 seconds, resulting in $N = 20160$.

Figure 7.2 shows how sample entropy of actigraphic data behaves with different parameter selection. The actigraphic recording shown in figure (A) in 3.1 was used to calculate the entropy values. We can see that the behaviour differs from that displayed in 7.1 in that the entropy grows with increasing r before achieving a peak and then falls as expected when r gets too large. This could be explained by the activity recorded during the night. Even though the tolerance r is small, the template vectors constructed using the data collected during sleep that are similar

for m points do remain similar for an additional point. This in effect results in a lower value of entropy. As the tolerance increases and the condition for template matching gets more relaxed, template matches of length m start occuring during the day. But because the activity while being awake is less predictable than that during sleep, these matched templates on the other hand do not remain similar for an additional point as often as those that occur during night.

m	A	B	entropy
2	26 330 538	32 753 064	0.218267
3	22 596 475	26 330 538	0.152935
4	19 979 358	22 596 475	0.123094
5	17 851 758	19 979 358	0.112598
6	16 084 232	17 851 758	0.104263

Table 7.1: The effect of embedding dimension m on the sample entropy value, computed with $r = 0.2$. A and B indicate the number of matched template pairs of length $m + 1$ and m, respectively.

Longer patterns in data can be looked for by increasing the embedding dimension m. This results in a more strict template matching and lower count of matched templates. It has been suggested that a recording of size $N = 10^m - 20^m$ is required in order to correctly estimate the conditional probabilities when computing the approximate entropy [34] but no recommendation for the minimal record length used with sample entropy has been proposed, and therefore this suggestion is sometimes respected even when computing the sample entropy. Nevertheless, the sample entropies with higher embedding dimension m in this thesis were computed even for the datasets where this condition did not hold true. Table 7.1 presents the number of matched template pairs when computing the entropy of recording (A) in 3.1. It indicates that if templates of higher length m do match, it is more probable that they will also match for an additional point.

7.2.3 Ability to distinguish signals of different complexity

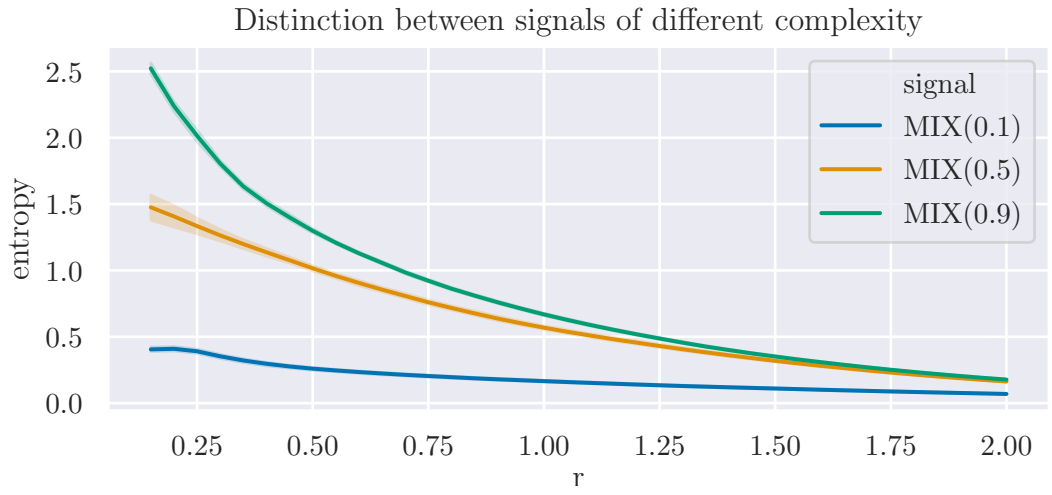

Figure 7.3: Ability of sample entropy to distinguish signals of different complexity. Computed with fixed $m = 2$ and variable r. Number of data points $N = 500$.

In the interest of providing an example of the discriminative ability of sample entropy, signals of varying complexity were simulated and their respective entropies were computed.

The simulated signal MIX(p) has been constructed as a mixture of a simple sinusoid, a highly regular signal, and white noise, an unpredictable source of data. The parameter p specifies the probability that a data point in the sinusoid is replaced by a randomly sampled value from the normal distribution. Figure 7.3 shows entropy values of the three generated signals, computed with fixed embedding dimension $m = 2$ and varying tolerance r. The replacement of the sinusoidal results in a worsened predictability of the time series, indicating a higher complexity. We can see that sample entropy successfully recognizes the different levels of complexity attained by each of these signals.

7.3 Handling of missing data

Several experiments were performed in order to assess the susceptibility of sample entropy to segments of missing values in the processed time series data. Utilization of three distinct techniques used to counter the error in the value of sample entropy was considered.

Linear interpolation Segments of missing data are filled linearly.

Skip method A segment of missing values is left out of the dataset and the two parts of the time series that were divided by this segment are connected.

Keep sample entropy A simple modification to the set of templates included during the computation of sample entropy was proposed in [35]. Whenever a template $\mathbf{x}_{m+1}(i)$ of length $m + 1$ contains a missing value, it is discarded from the entropy computation along with its corresponding template $\mathbf{x}_m(i)$ of length m. By doing so, it is prevented that the measurements that could have been far apart from each other in the original dataset are associated with each other to form a new template vector, as they would have if for example the skip method had been used.

As during the investigation of the effect of missing data on the values of nonparametric analysis variables, both real-world datasets and generated datasets were used. The method of simulating missing segments of data remained the same as well. Each percentage of missing data was simulated 300 times.

7.3.1 Experiment results

Figure 7.4 shows the experiment results where random data segments were removed from the real data. It is not surprising that the linear interpolation of missing data has the worst performance out of all the considered techniques. It introduces new data that is highly regular, which subsequently underestimates the real value of signal complexity.

It can be seen that the skip method and keep sample entropy achieve virtually the same level of error. A more detailed comparison of these two is shown in figure 7.5. It is also apparent that both methods are more resilient to shorter, although more frequent, missing segments of data.

The same experiment conducted using artificially generated data confirms the previous findings, as can be seen in figure 7.6. Keep sample entropy and the skip method performed equally well. The response to the number of removed continuous parts of data is also very much like in the previous case, when the real data had been used. However, the maximum level of error is lower than previously estimated.

In conclusion, the skip method has been used with recordings containing missing values.

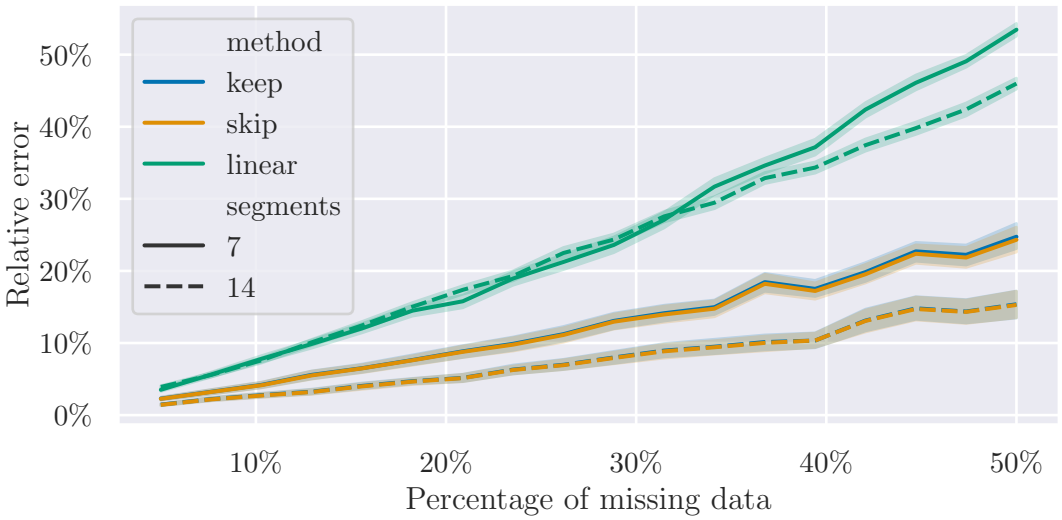

Figure 7.4: Effect of missing data on the value of sample entropy. Simulated using real data.

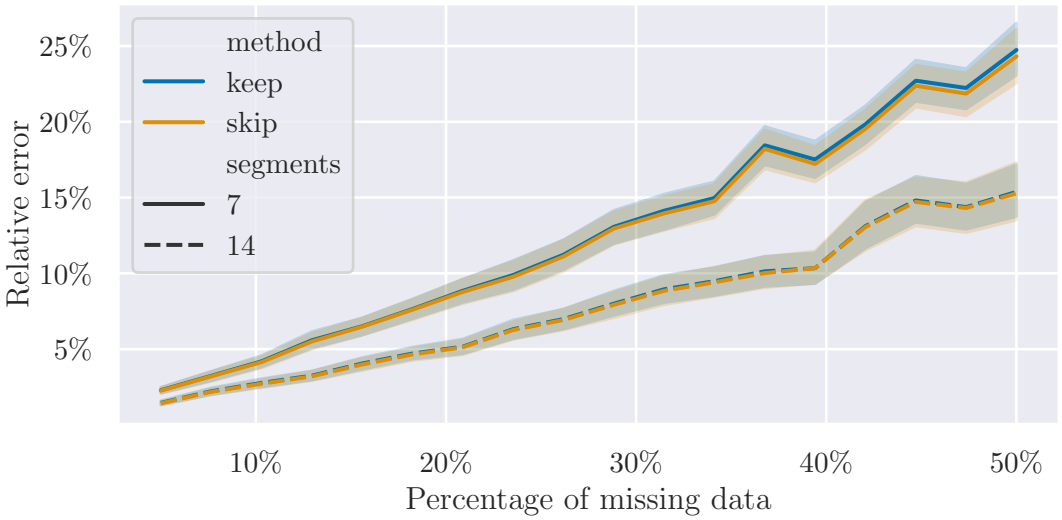

Figure 7.5: Effect of missing values on the value of sample entropy of real-world data, showing the performances of the skip method and keep sample entropy in greater detail.

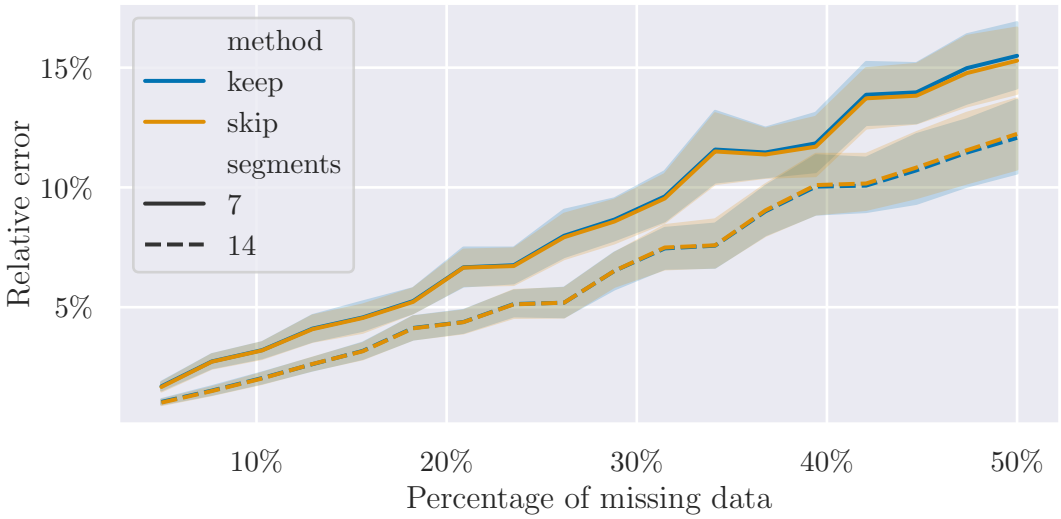

Figure 7.6: Effect of missing data on the value of sample entropy. Simulated using artificial data.

Chapter 8

Data Analysis

8.1 Sample entropy grid search

Based on the description and characteristics of episode types in bipolar disorder, it can be expected that the complexity of motor activity will differ between depression, mania and euthymia. In comparison with the entropy values observed during euthymia, the entropy values could attain higher levels during mania due to its impulsive and chaotic nature, and lower levels during episodes of depression due to more frequent periods of inactivity.

A grid search has been conducted in order to find entropy parameters that would be most suitable for capturing the change in complexity of physical activity with respect to the current mood state of the subject. The data of 12 patients obtained as part of the longitudinal study were used for this purpose and the submitted questionnaires were referenced as the, although subjective, measurements of mood state in order to assess the quality of an entropy setting.

Apart from the tolerance r and the embedding dimension m, the length of the actigraphic recording and its sampling period were considered as parameters as well. A number of recordings of different durations were extracted for each of the questionnaire submissions. The recording always included the day of the submission and an additiononal number of days prior to it, so that the total length of the recording would equal the recording length parameter as specified by grid search.

The questionnaire submissions were split into a training set and a test set, so that the performance of the configurations could have been validated. The split was done so that 70% of each patient's submissions were included in the training set and the rest was included as a part of the test set.

The considered values for the embedding dimension m were 2, 3, 4, 5, and 6, the values of tolerance r started at 0.15 and then went from 0.2 up to 0.9 in increments of 0.1. The lengths of the extracted recordings were 2 to 7 days and 10, 12, and 14 days. The examined sampling periods were 30 seconds and 2, 5, 10, 15, 20, and 30 minutes for the shorter durations up to and including 7 days. However, the sampling periods considered for the recordings of 10, 12, and 14 days did not include 30 seconds and were expanded by the sampling of 1 hour. Recordings missing more than 10% of data were not included in the grid search so that the errors in entropy values would not be too significant.

The performance of an entropy setting was measured as follows. The sample entropies were computed for all the extracted actigraphy data using the considered setting. After that, the Spearman's rank correlation coefficients between the entropies and their corresponding submission's sum of marks assigned to the statements associated with depression and the sum of marks assigned to the statements associated with mania were computed for each of the patients individually.

These correlations have then been combined and treated as vectors

$$\begin{pmatrix} \rho_{p,d} \\ \rho_{p,m} \end{pmatrix} \in \mathbb{R}^2$$

where $\rho_{p,d}$ is the correlation of the entropy and the sums of depression marks of the patient p and $\rho_{p,m}$ is the correlation of the entropy and the sums of mania marks of the patient p.

Finally, the performance score of the inspected entropy setting was measured as the mean square of distances between these vectors and the vector of the desired correlation coefficients

$$\frac{1}{|\mathbb{P}|}\sum_{p\in\mathbb{P}}\left\|\begin{pmatrix} \rho_{p,d} \\ \rho_{p,m} \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\|^2 \tag{8.1}$$

where $\mathbb{P}$ is the set of patients included in the grid search. Lower resulting values of this measure indicate a better entropy performance. As all the patients contributed a single measurement towards the entropy setting's final measure of quality, no bias was introduced even though the number of questionnaire submissions was not balanced patient-wise.

8.1.1 Results

The lowest value of the measure 8.1 was achieved by extracting recordings that have a duration of 6 days and 30 second sampling period, and setting the embedding dimension $m = 2$ and the tolerance $r = 0.2$. Note that the sample entropy parameters are in the range of the original suggestions for the parameters m and r made by Pincus et. al [30].

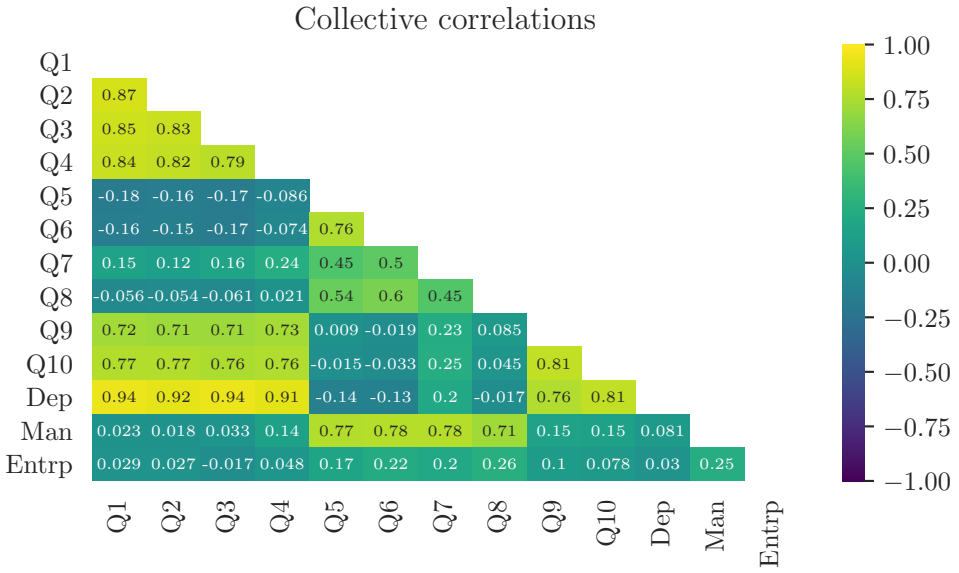

Figure 8.1: Collective correlations of entropy and questionnaire submissions. Entropy was computed using the found setting. The variables Dep and Man stand for the sum of marks assigned to the statements meant to assess symptoms of depression and mania, respectively.

Figure 8.1 shows the correlations between the questionnaire submissions and the sample entropy computed using the found setting for all the patients whose data was included in the grid search. It can be seen that even though this setting should have the best performance out of those that were considered, its resulting entropies do not anticorrelate with the sum of marks assigned to the statements associated with depression as expected. On the other hand, a low value of positive correlation with the mania statement marks sums is present.

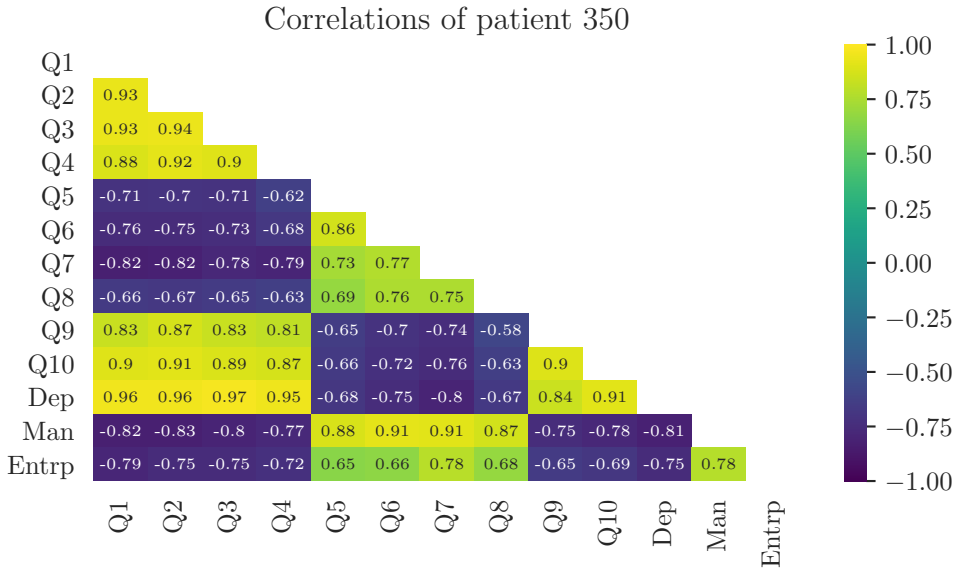

Figure 8.2: Correlations between questionnaire submissions and entropy for patient 350.

However, more promising results can be found if the correlations are investigated individually for each of the patients. For example, the entropy values achieve a strong correlation with the survey statements for the patient 350, as presented in figure 8.2, and a moderate strength for patient 928, as seen in figure 8.3.

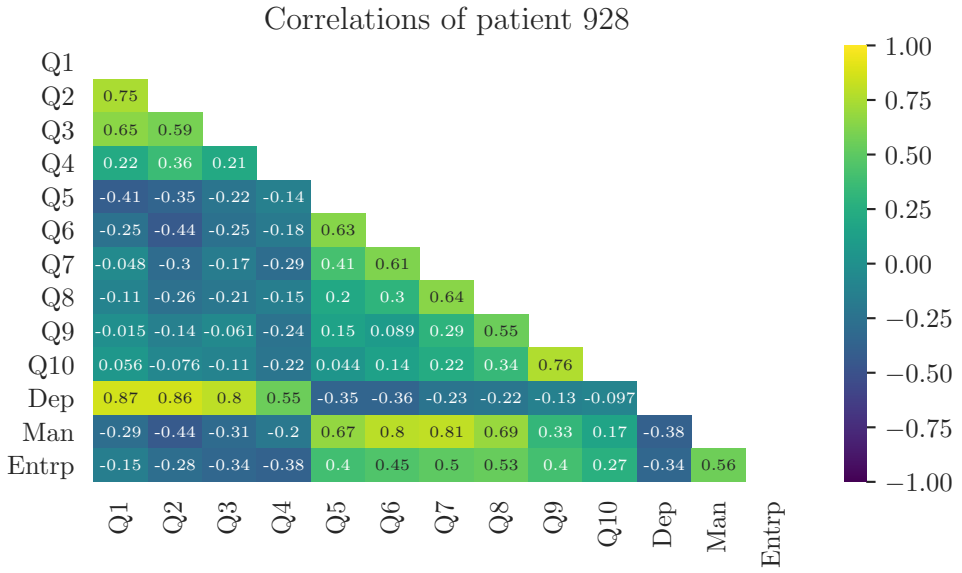

Figure 8.3: Correlations between questionnaire submissions and entropy for patient 928.

Nevertheless, the setting did not work correctly in all the inspected patients. In some cases the correlations were very low or none at all. For example, the correlations of patient 574 shown in figure 8.4, where the correlations are very low and, moreover, the entropy is unexpectedly correlated with depression.

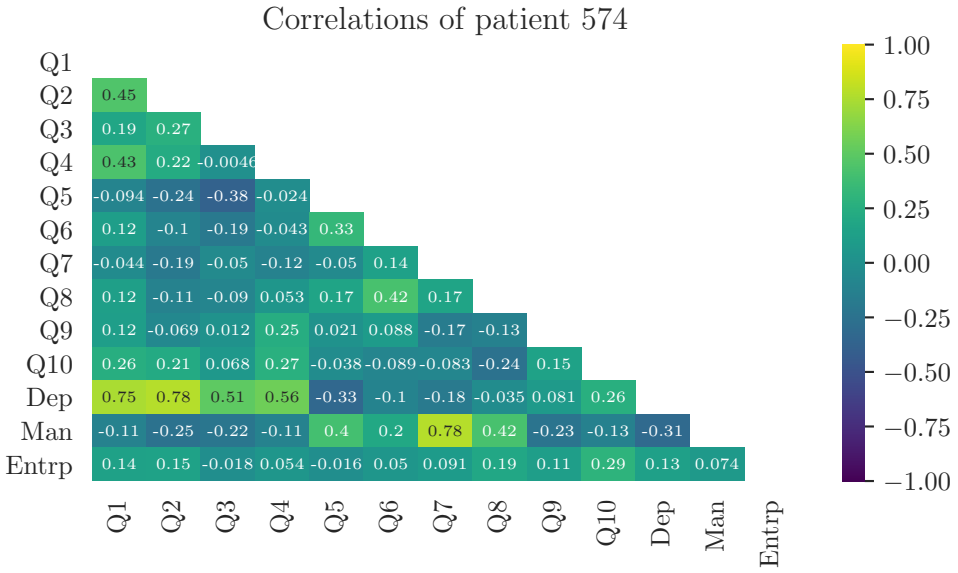

Figure 8.4: Correlations between questionnaire submissions and entropy for patient 574.

8.2 Nonparametric analysis

The behaviour of nonparametric variables during individual episodes and remission was investigated. For this purpose, multiple non-overlaping segments with duration of one week were extracted from the hand-labeled data segments. The number of records contributed by individual patients was balanced for each of the three states of mania, depression, and remission. Therefore no bias towards a specific individual was expected. The sets of patients whose data was used were disjunct between the states, with an exception of two patients' data that was included in both the depression and mania datasets.

8.2.1 IV and IS values during episodes

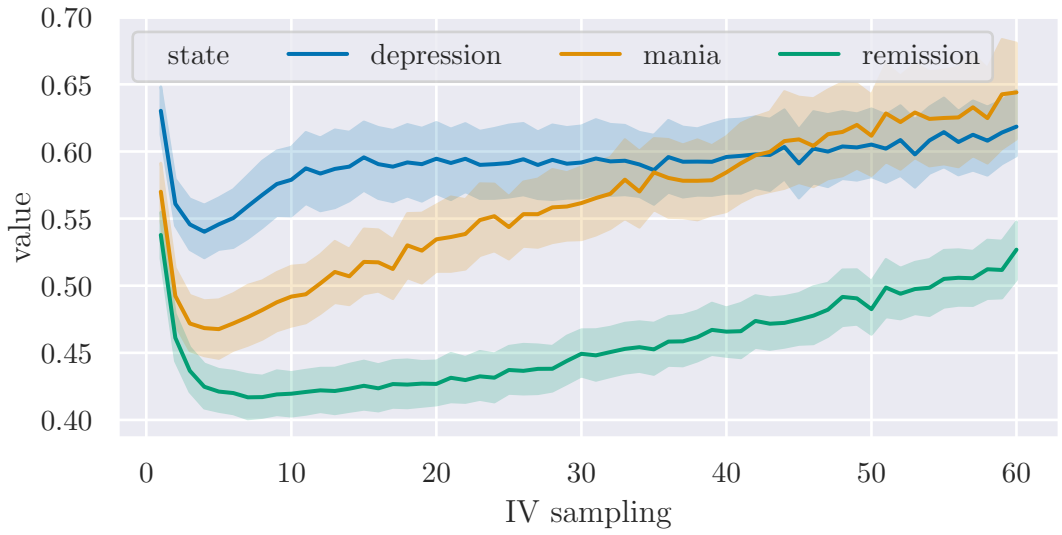

Figure 8.5: Differences in IV between individual states.

Figure 8.5 displays the difference of the IV variables between the mood states. The horizontal axis displays the minute sampling used to compute the variable. There is an apparent distinction between the episodes using the IV10 variable, with depression achieving the largest value, remission the lowest value and mania being in the middle. This result is in accordance with the expectation that IV attains larger values when frequent transitions between activity and idleness occur in the actigraphic recording and therefore should be higher during depression, as inactivity during the day is characteristic for this type of episode. Although as the rate of sampling used to compute the variable grows, the difference between depression and mania fades away. Nevertheless, the distinction between remission and both types of episodes is still discernible.

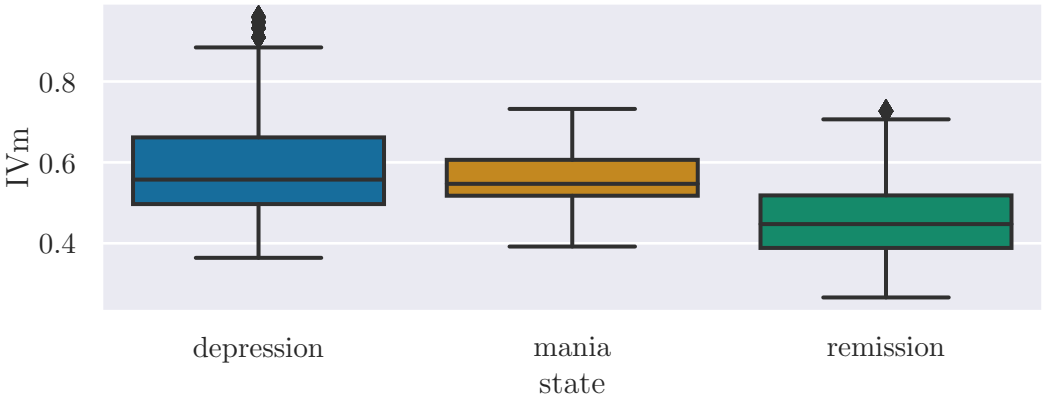

Figure 8.6: Comparison of mood states using the IVm variable.

The values of IVm for each of the conditions are summarized in figure 8.6. As expected by inspecting the previous graph of IV values, there is a difference between remission and both the depression and mania. However, the distinction between depression and mania is more pronounced in the IV10 variable, shown in figure 8.7.

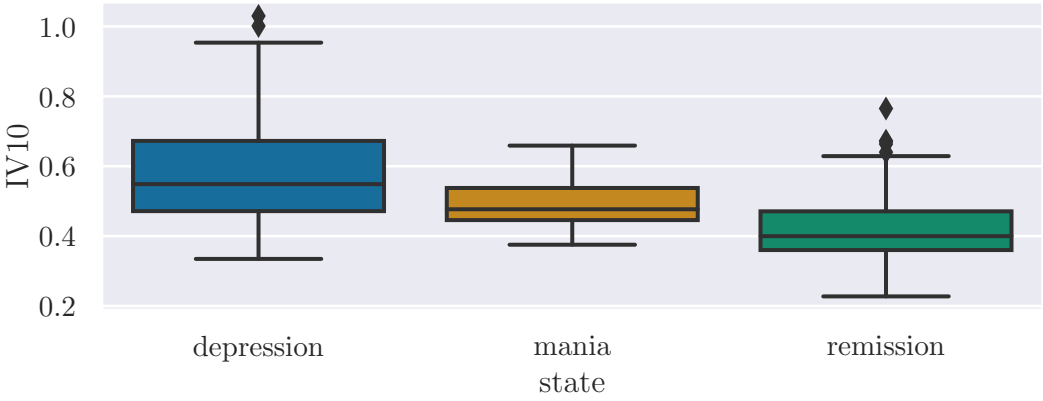

Figure 8.7: More pronounced difference between depression and mania found using the IV10 variable.

The graph displayed in figure 8.8 shows the values of IS variables for each of the groups. It can been seen that according to the interdaily stability, the circadian rhythm is most stable during periods of remission. The stability is impaired during episodes of depression and mania, but there is almost no difference in stability between these two episode types. This indicates that the values of IS could help discern between a bipolar episode and a remission, but not between the episode types themselves. The interdaily stability mean is shown in figure 8.9.

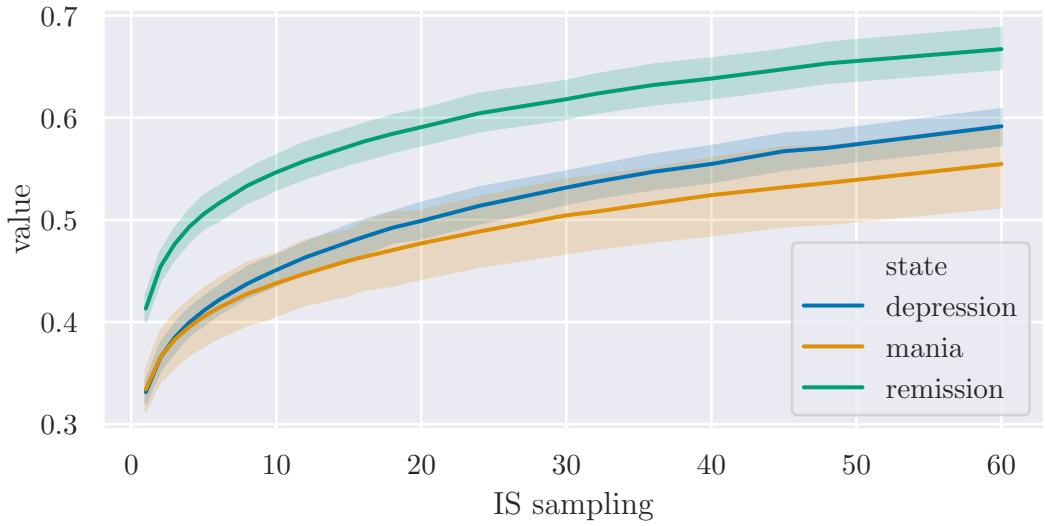

Figure 8.8: Differences in IS between individual states.

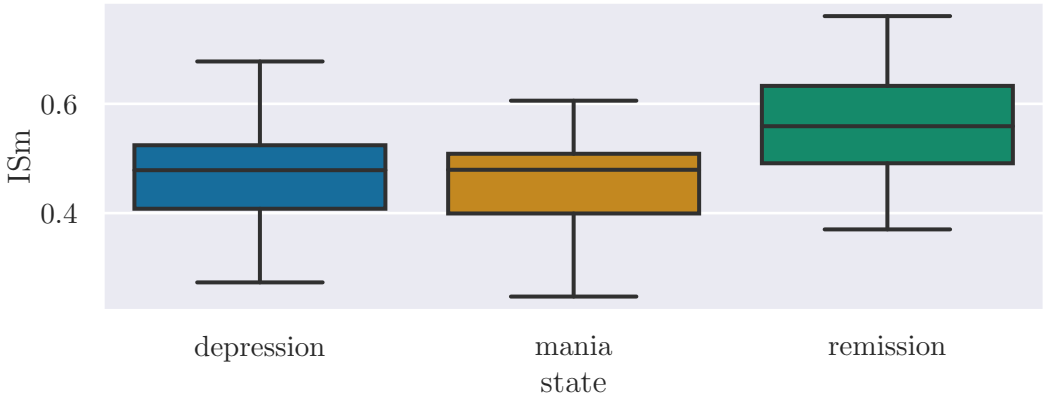

Figure 8.9: Comparison of mood states using the ISm variable.

8.2.2 M10 and L5 values during episodes

Figure 8.10 presents the values of the average activity during the most active 10 hours. The values of M10 observed during episodes of depression conform with the expectation of lower overall activity during this episode type, as they achieve lower values than during episodes of mania or remission.

From figure 8.11 it is apparent that the values of L5 are more likely to achieve higher values during episodes of mania than during depression or remission. This can be attributed to a lesser need to sleep, which is one of the symptoms of mania. Therefore the values of L5 could be of help when detecting this type of episode.

Overall, subtle differences between the variables of nonparametric analysis have been found during each of the episode types and remission, but none of the variables displayed an ability to clearly differentiate between these states.

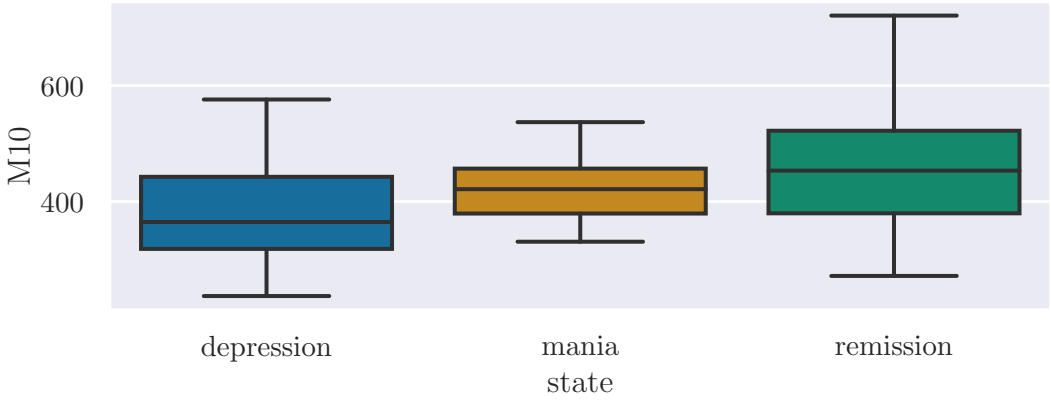

Figure 8.10: The values of M10 during each of the episode types and remission.

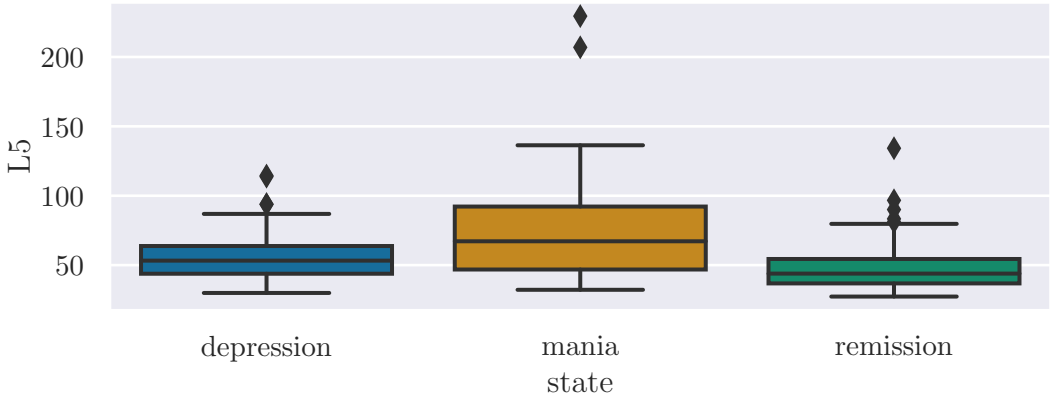

Figure 8.11: Comparison of the L5 variable during each of the states.

8.2.3 Nonparametric variables correlation

The correlation between the nonparametric variables was investigated as well. The Spearman's rank correlation coefficient ρ was used, as the relationship between individual variables was not expected to be linear. I randomly selected 10 week-long periods from 14 patients, so that the obtained data would be balanced. The resulting correlations are displayed in table 8.1.

	IVm	ISm	M10	L5	RA
IVm	-				
ISm	-0.53	-			
M10	-0.80	0.57	-		
L5	0.18	-0.53	-0.06	-	
RA	-0.49	0.73	0.43	-0.91	-

Table 8.1: Correlations between individual nonparametric variables.

As can be expected, the relative amplitude is correlated with M10 and anticorrelated with L5. However, the relationship of RA with L5 is much stronger than that with M10. This can be easily explained by examining formula 6.3 used to compute RA. An increase in M10 increases both the numerator and the denominator, whereas an increase in L5 decreases the value of the numerator and increases the value of the denominator, which results into a significantly greater change in the value of RA and consequently has a larger effect on the ordering of RA values.

Another interesting outcome is the anticorrelation between M10 and IVm. M10 achieves higher values if the level of activity during the whole period of 10 hours from which it is computed does not significantly fall, which would lower the average activity and therefore the M10 variable. However, as the fluctuations in activity do not occur, the IVm attains a low value.

Chapter 9

Patient state classificator

Now that the methods have been presented and used for inspecting the differences between the individual episodes of bipolar disorder, it can be examined how well they can be made use of for the purpose of predicting the state of the patient based on the recorded physical activity.

Three classificators were constructed in total, with each of them differentiating between a different pair of mood states.

9.1 Datasets construction

The hand-labeled data was used while training and testing the machine learning models. All the recorded episodes were split into week-long segments and subsequently their features were computed. These segments inherited the label of the episode from which they originated.

Using the same set of data and procedure as during the grid search in section 8.1, it was found that the optimal setting for computing sample entropy of a record that consists of 7 days does not differ from that found previously. Therefore the entropy of each of the extracted actigraphy recordings was computed using their original sampling period of 30 seconds and using the embedding dimension $m = 2$ and tolerance $r = 0.2$.

The data was then split into two seperate datasets, a training dataset and a test dataset. Each variable was standardized to have a 0 mean and a unit variance. The values used for standardization were inferred from the training set. Two main precautions must have been taken while constructing these datasets, so that the resulting performance of the classifiers could be considered to be valid.

Firstly, the classes—mood states—present in the datasets were balanced. This was achieved by undersampling the classes with higher number of records. By doing so, it was prevented that the machine learning models would simply infer the probability distribution of the more populated class and develop a bias towards predicting its label. As a consequence, the datasets containing the mania class contained less entries than other datasets, because mania was the least prevalent mood state in the provided data and also because manic episodes are shorter in duration than depressive episodes, which resulted into fewer entries.

Secondly, the number of entries contributed by each patient towards a dataset was not balanced between all the included patients, and therefore one might propose that the classifiers did not learn how to recognize the differences between individual mood states, but learned to recognize patients that contributed the most entries. However, the sets of patients included in each training dataset and its respective test dataset were disjunct and no patient's entries were present in both of them. By doing so, the ability of the model to predict a patient's state was tested as opposed to testing the ability to recognize a patient.

The training and testing datasets were constructed as described for each of the three classificators.

The considered machine learning models were support vector machine, logistic regression, decision trees and random forests. However, only the best performing model is presented for each of the state pairs. The features used for training each of the models were sample entropy, M10, L5, IV and IS. The sampling periods using which the IV and IS variables were computed were selected with respect to the pair of states that were being classified based on the previously made analysis. A dimensionality reduction using the principal component analysis (PCA) was applied where it was found to be beneficial. The reported accuracies and other performance measures were assessed using the test datasets.

9.2 Results

The performance of the constructed classificators is reported in the following subsections. The following performanace measures were used to assess the quality of the classificators.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9.1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9.2}$$

where TP is the number of correctly classified instances of the positive class, FP is the number of misclassified instances of the negative class, and FN is the number of misclassified instances of the positive class.

These measures can be further combined to form the performance measure F1

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9.3}$$

9.2.1 Remission-Depression classificator

The best performance in distinguishing remission between depression was achieved by using a logistic regression, which displayed an accuracy of 63%. The IV20 and IS20 variables were selected, leveraging the analysis performed in section 8.2 of the previous chapter.

State	Precision	Recall	F1
Remission	0.61	0.75	0.67
Depression	0.67	0.52	0.59

Table 9.1: Performance measures of the remission-depression classificator. Only 52% of the depressive records are recognized and correctly classified as opposed to 75% of the records of remission.

Variable	Entropy	M10	L5	IS20	IV20
Coefficient	-0.22	0.1	-0.11	-0.36	0.43

Table 9.2: Resulting coefficients of the logistic regression that differentiates between remission and depression, where depression was treated as the positive class. These results are in accordance with the expectations of how these variables behave during each of the two states. The variable M10 is the only exception as the classifier expects it to be larger during the episodes of depression than during remission.

Table 9.2 shows the resulting coefficients of the logistic regression. As expected, a less stable circadian rhythm and a more fragmented activity were associated with the episodes of depression, as expressed by the coefficients of the IS20 and IV20 variables. Lower values of entropy were

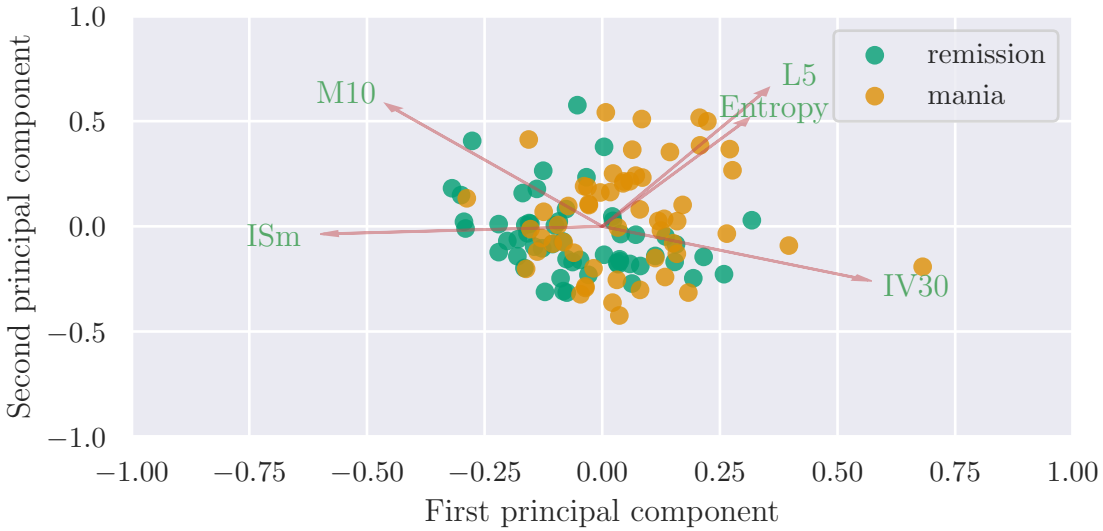

Figure 9.1: The biplot of the training dataset for the remission-mania classificator transformed using the PCA. It can be seen that the episodes of mania are associated with larger values of L5 and entropy. The remission records seem to be characterized by more stable and less fragmented circadian rhythm.

also recognized to differentiate depression from remission. However, the classificator surprisingly expects the variable M10 to be higher during depression than during remission.

9.2.2 Remission-Mania classificator

An accuracy of 70% was achieved by using a support vector machine with a radial basis function (RBF) kernel and by reducing the dimensionality of the data to 2 using the PCA. The hyperparameter γ of the RBF kernel was selected using a cross-validated grid search, which utilized the training data. The best found value was $\gamma = 0.65$. The interdaily stability mean was considered in this case and the intradaily variability values were computed using a 30-minute sampling period.

State	Precision	Recall	F1
Remission	0.69	0.64	0.67
Mania	0.67	0.71	0.69

Table 9.3: Performance measures of the remission-mania classificator. The achieved precision is similar for both of the classes, however the recall is higher for the manic class.

The resulting training dataset after dimensionality reduction is displayed in figure 9.1. According to this biplot it seems that the episodes of mania achieved larger values of sample entropy and L5 and show a less stable circadian rhythm than the records of remission. Figure 9.2 shows the fitted decision boundary of the SVM and the data points of training dataset and the testing dataset. It is apparent that the classificator's decision boundary encompasses an area that excludes the entries with large values of L5 and entropy. The entries inside of this area get classified as remissions while the entries that fall outside of this area are classified as manias.

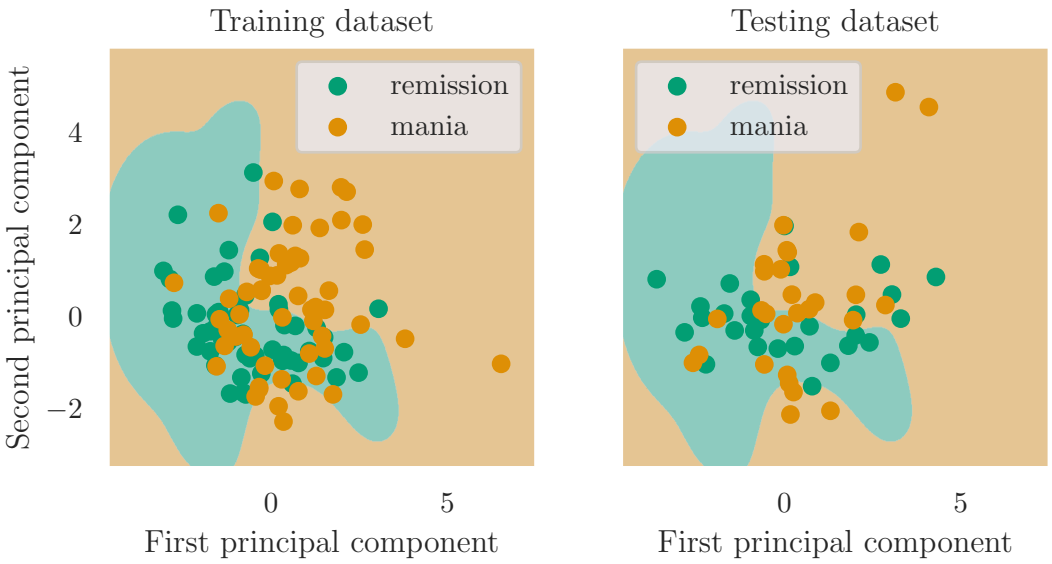

Figure 9.2: The decision boundary of the SVM classificator that differentiates remission from mania. The boundary is shown along with the training data on the left-hand side and the testing data on the right-hand side.

9.2.3 Depression-Mania classificator

The accuracy of determining whether an activity was recorded during an episode of depression or during an episode of mania was 72%. It was achieved using a support vector machine with the RBF kernel. Once again, the hyperparameter γ was tuned using a cross-validated grid search, which yielded a value of $\gamma = 0.326$. The ISm and IV10 values were selected to be used for training this classificator. The dimensionality of data was not reduced as it was not found to be helpful, and therefore the decision boundary could not have been displayed. This model included the starting times of the periods from which the M10 and L5 values were computed as additional features.

State	Precision	Recall	F1
Depression	0.70	0.76	0.73
Mania	0.74	0.68	0.71

Table 9.4: Performance measures of the depression-mania classificator. The recognition of mania is more precise than the recognition of depression, however the recall of depression is better than that of mania.

Table 9.4 shows the performance of the classificator for each of the two classes. The precisions of their recognition are similar but the precision of recognizing mania is slightly better, while the depression achieves a better recall value.

Chapter 10

Conclusion

This thesis investigated the differences in motor activity recorded using an actigraphic device among different mood states in patients with bipolar disorder. Analyzing the recorded activity could help recognize a change in a patient's mood state as well as the type of the occuring episode. This could then be leveraged by health care professionals and enable them to provide quick assistance and treatment during an acute episode.

In the first part of this thesis, an overview of bipolar disorder and of the method used to capture motor activity was given. Two techniques used to analyze the recorded activity were presented: the nonparametric actigraphy analysis, allowing the study of the circadian activity rhythm, and sample entropy, a measure quantifying the amount of regularity in the recorded data. The effect of missing data on the variables used in these methods was investigated. The techniques used to handle the presence of missing data during the computation of sample entropy were examined as well.

Grid search was conducted in order to find a setting of sample entropy and an optimal duration as well as the sampling of the actigraphic record that, when combined, would be the most suitable for capturing the change in regularity of a patient's motor activity with respect to his or her current mood state. The best results were achieved when 6-day-long recordings captured using a 30-second sampling period were analyzed using sample entropy with the tolerance $r = 0.2$ and the embedding dimension $m = 2$. This setting also displayed the best performance when analyzing records with the duration of one week. The sample entropy attained lower values during episodes of depression and higher values during episodes of mania, although this result was not observed in all of the patients. Future work could examine a more personalized approach towards estimating the regularity of physical activity.

The nonparametric analysis of actigraphy data was used to analyze the difference in activity during different types of episodes and euthymia. Subtle differences were found between the mood states in all of the inspected variables, with the exception of IS values that did not differ between individual bipolar episodes but differed between the episodes and euthymia. These findings were in agreement with the previous expectactions about the behaviour of these variables during each of the episode types.

Finally, three classificators were constructed. Each of them was used to differentiate between two different episode types. These classificators utilized the information extracted using the described activity analysis techniques. The best performance with an accuracy of 72% was achieved when distinguishing between depression and mania, followed by an accuracy of 70% when discriminating between remission and mania. The worst performing classificator, which separated remission and depression, achieved an accuracy of 63%, with the recall of depression being just 52%. Even though these classificators could provide useful results, they are not fit to be used in real-world setting. However, they could be further expanded to include features computed using different methods of activity analysis, for example cosinor analysis, in an attempt to improve their usability and reliability.

Bibliography

[1] K. R. Merikangas, R. Jin, J.-P. He, R. C. Kessler, S. Lee, N. A. Sampson, M. C. Viana, L. H. Andrade, C. Hu, E. G. Karam, M. Ladea, M. E. Medina-Mora, Y. Ono, J. Posada-Villa, R. Sagar, J. E. Wells, and Z. Zarkov, "Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative", *Archives of General Psychiatry*, vol. 68, no. 3, pp. 241–251, Mar. 2011, ISSN: 0003-990X. DOI: 10.1001/archgenpsychiatry.2011.12. eprint: https://jamanetwork.com/journals/jamapsychiatry/articlepdf/211192/yoa05088_241_251.pdf. [Online]. Available: https://doi.org/10.1001/archgenpsychiatry.2011.12.

[2] A Martinez-Aran, E Vieta, C Torrent, J Sanchez-Moreno, J. Goikolea, M Salamero, G. Malhi, A Gonzalez-Pinto, C Daban, S Alvarez-Grandi, *et al.*, "Functional outcome in bipolar disorder: The role of clinical and cognitive factors", *Bipolar disorders*, vol. 9, no. 1-2, pp. 103–113, 2007.

[3] J. Alonso, M. Petukhova, G. Vilagut, S. Chatterji, S. Heeringa, T. B. Üstün, A. O. Alhamzawi, M. Viana, M. Angermeyer, E. Bromet, *et al.*, "Days out of role due to common physical and mental conditions: Results from the who world mental health surveys", *Molecular psychiatry*, vol. 16, no. 12, pp. 1234–1246, 2011.

[4] M. Pompili, X. Gonda, G. Serafini, M. Innamorati, L. Sher, M. Amore, Z. Rihmer, and P. Girardi, "Epidemiology of suicide in bipolar disorders: A systematic review of the literature", *Bipolar disorders*, vol. 15, no. 5, pp. 457–490, 2013.

[5] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[6] L. Forty, D. Smith, L. Jones, I. Jones, S. Caesar, C. Cooper, C. Fraser, K. Gordon-Smith, S. Hyde, A. Farmer, and et al., "Clinical differences between bipolar and unipolar depression", *British Journal of Psychiatry*, vol. 192, no. 5, 388–389, 2008. DOI: 10.1192/bjp.bp.107.045294.

[7] I. Grande, M. Berk, B. Birmaher, and E. Vieta, "Bipolar disorder", *The Lancet*, vol. 387, no. 10027, pp. 1561–1572, 2016. DOI: 10.1016/S0140-6736(15)00241-X. [Online]. Available: https://doi.org/10.1016/S0140-6736(15)00241-X.

[8] E. Vieta and T. Suppes, "Bipolar ii disorder: Arguments for and against a distinct diagnostic entity", *Bipolar Disorders*, vol. 10, no. 1p2, pp. 163–178, 2008.

[9] B. N. Frey, A. C. Andreazza, J. Houenou, S. Jamain, B. I. Goldstein, M. A. Frye, M. Leboyer, M. Berk, G. S. Malhi, C. Lopez-Jaramillo, *et al.*, "Biomarkers in bipolar disorder: A positional paper from the international society for bipolar disorders biomarkers task force", *Australian & New Zealand Journal of Psychiatry*, vol. 47, no. 4, pp. 321–332, 2013.

[10] R. C. Young, J. T. Biggs, V. E. Ziegler, and D. A. Meyer, "A rating scale for mania: Reliability, validity and sensitivity", *The British journal of psychiatry*, vol. 133, no. 5, pp. 429–435, 1978.

[11] S. A. Montgomery and M. Åsberg, "A new depression scale designed to be sensitive to change", *The British journal of psychiatry*, vol. 134, no. 4, pp. 382–389, 1979.

[12] M. Hamilton, "A rating scale for depression", *Journal of neurology, neurosurgery, and psychiatry*, vol. 23, no. 1, p. 56, 1960.

[13] M. L. Phillips and D. J. Kupfer, "Bipolar disorder diagnosis: Challenges and future directions", *The Lancet*, vol. 381, no. 9878, pp. 1663–1671, 2013.

[14] J. R. Geddes and D. J. Miklowitz, "Treatment of bipolar disorder", *The lancet*, vol. 381, no. 9878, pp. 1672–1682, 2013.

[15] R. Belmaker, "Bipolar disorder", *New England Journal of Medicine*, vol. 351, no. 5, pp. 476–486, 2004.

[16] J. L. Schulman and J. M. Reisman, "An objective measure of hyperactivity", *Am J Ment Defic*, vol. 64, pp. 455–456, Nov. 1959.

[17] C. F. Johnson, "Hyperactivity and the machine: The actometer", *Child Development*, vol. 42, no. 6, pp. 2105–2110, 1971, ISSN: 00093920, 14678624. [Online]. Available: `http://www.jstor.org/stable/1127617`.

[18] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms", *Sleep*, vol. 26, no. 3, pp. 342–392, 2003.

[19] S. H. Jones, D. J. Hare, and K. Evershed, "Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder", *Bipolar disorders*, vol. 7, no. 2, pp. 176–186, 2005.

[20] K. Krane-Gartiser, T. E. G. Henriksen, G. Morken, A. Vaaler, and O. B. Fasmer, "Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder", *PloS one*, vol. 9, no. 2, 2014.

[21] L. Tondo, G. H Vazquez, and R. J Baldessarini, "Depression and mania in bipolar disorder", *Current neuropharmacology*, vol. 15, no. 3, pp. 353–358, 2017.

[22] A. G. Harvey, "Sleep and circadian functioning: Critical mechanisms in the mood disorders?", *Annual review of clinical psychology*, vol. 7, pp. 297–319, 2011.

[23] W. Witting, I. Kwa, P. Eikelenboom, M. Mirmiran, and D. Swaab, "Alterations in the circadian rest-activity rhythm in aging and alzheimer's disease", *Biological Psychiatry*, vol. 27, no. 6, pp. 563 –572, 1990, ISSN: 0006-3223. DOI: `10.1016/0006-3223(90)90523-5`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/0006322390905235`.

[24] B. S. Gonçalves, P. R. Cavalcanti, G. R. Tavares, T. F. Campos, and J. F. Araujo, "Nonparametric methods in actigraphy: An update", *Sleep Science*, vol. 7, no. 3, pp. 158 – 164, 2014, ISSN: 1984-0063. DOI: `10.1016/j.slsci.2014.09.013`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1984006314000510`.

[25] Y.-L. Huang, R.-Y. Liu, Q.-S. Wang, E. J. Van Someren, H. Xu, and J.-N. Zhou, "Age-associated difference in circadian sleep–wake and rest–activity rhythms", *Physiology & behavior*, vol. 76, no. 4-5, pp. 597–603, 2002.

[26] E. J. Van Someren, D. F. Swaab, C. C. Colenda, W. Cohen, W. V. McCall, and P. B. Rosenquist, "Bright light therapy: Improved sensitivity to its effects on rest-activity rhythms in alzheimer patients by application of nonparametric methods", *Chronobiology international*, vol. 16, no. 4, pp. 505–518, 1999.

[27] G. Cornelissen, "Cosinor-based rhythmometry", *Theoretical Biology and Medical Modelling*, vol. 11, no. 1, p. 16, 2014.

[28] J. Wang, H. Xian, A. Licis, E. Deych, J. Ding, J. McLeland, C. Toedebusch, T. Li, S. Duntley, and W. Shannon, "Measuring the impact of apnea and obesity on circadian activity patterns using functional linear modeling of actigraphy data", *Journal of circadian rhythms*, vol. 9, no. 1, p. 11, 2011.

[29] C. E. Shannon, "A mathematical theory of communication", *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[30] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis", *Journal of clinical monitoring*, vol. 7, no. 4, pp. 335–345, 1991.

[31] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy", *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, H2039–H2049, 2000.

[32] D Abásolo, R Hornero, P Espino, D Alvarez, and J Poza, "Entropy analysis of the eeg background activity in alzheimer's disease patients", *Physiological measurement*, vol. 27, no. 3, p. 241, 2006.

[33] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability", *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, R789–R797, 2002.

[34] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: What does regularity quantify?", *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 266, no. 4, H1643–H1656, 1994.

[35] X. Dong, C. Chen, Q. Geng, Z. Cao, X. Chen, J. Lin, Y. Jin, Z. Zhang, Y. Shi, and X. D. Zhang, "An improved method of handling missing values in the analysis of sample entropy for continuous monitoring of physiological signals", *Entropy*, vol. 21, no. 3, p. 274, 2019.