INAUGURAL-DISSERTATION

zur Erlangung der Doktorwürde
der Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

B.Sc. Informatikerin, M.Sc. Angewandte Mathematikerin
Lilian Villarín Pildaín
aus
Havanna/ Kuba

Tag der mündlichen Prüfung:_____

Thema:

Alcohol drinking pattern analysis - an in silico tool to model and predict addictive behaviors

Gutachter: **Prof. Dr. Willi Jäger**

A mis padres.

Acknowledgements

Abstract

We present methods for the systematic modelling and clustering of time series. Our data is associated with behavioral studies of alcoholism in animals. We analyze multivariate time series obtained from an automated *drinkometer* system. Here, rats have free access to water and three alcoholic solutions (this being the baseline treatment level), which is then interrupted by repeated deprivation phases. We develop a methodology to simultaneously classify into- and characterize dynamic patterns of the observed drinking behavior. This is achieved by extending known results on *generalized linear models (GLM)* for univariate time series to the multivariate case. We simplify the computational fitting procedure, by assuming a shared seasonal pattern throughout individuals and implementing an *estimation maximization (EM)* algorithm to fit mixtures of the mentioned multivariate GLM. A partition of the data, as well as a characterization of each group is obtained. The observed patterns of drinking behavior differ in their preference profile for the three alcoholic solutions, and also in the net alcohol intake. We observe an evolution of the drinking behavior over the repeated cycles of alcohol admission and deprivation, with a clear initial preference profile and a development to one of the advanced profiles. Furthermore, to measure the alcohol deprivation effect in this 4-bottle setting, a new criterion is developed, which enables us to classify each rat into presenting ADE or not. This classification shows that the rats develop a tolerance to taste adulteration after few deprivation phases. The proposed framework can be employed to find differences in behavior between different conditions and/or groups of animals and in the prediction of alcoholism from early phases of alcohol intake. The developed methods can also be used in different fields, where the analysis of time series plays an important role (e.g. microarray analysis and neuroscience).

Zusammenfassung

In dieser Arbeit werden Methoden für die Modellierung und das Clustering von Zeitreihen entwickelt und angewandt, um die Entwicklung von Alkoholismus im Tiermodell zu beschreiben. Die erhobenen Daten sind hierbei multivariate Zeitreihen, die mittels eines automatisierten Gerätes, des *Drinkometers*, erhoben werden. Im Experiment haben Ratten freien Zugang zu Wasser und drei alkoholischen Lösungen („Baseline"-Trinkverhalten), zyklisch unterbrochen von „Entzugsphasen", nach denen ein „Alkohol-Deprivations-Effekt" (ADE) beobachtet wird. Zur Messung des ADEs wurde ein neues Kriterium entwickelt, welches eine Klassifizierung jedes einzelnen Individuums bezüglich des ADEs erlaubt. Damit kann gezeigt werden, dass Ratten eine Toleranz gegenüber vergällten alkoholischen Lösungen nach mehreren Entzugsphasen entwickeln. Die in dieser Arbeit entwickelten Methoden klassifizieren und charakterisieren dynamische Muster des beobachteten Alkohol-Trinkverhaltens. Der Hauptansatz ist dabei die Erweiterung verallgemeinerter linearer Modelle (GLM) für univariate Zeitreihen auf den multivariaten Fall. Das Fitting-Verfahren wurde durch die Annahme zirkadianer Komponenten vereinfacht, sowie durch die Implementierung eines EM-Algorithmus, um die Modelle der multivariaten GLM zu fitten. Dadurch wurde sowohl eine Klassifizierung der Trinkmuster in verschiedene Cluster, als auch eine Charakterisierung der Dynamik dieser Muster erreicht. Diese unterscheiden sich in ihrem Präferenz-Profil für die drei alkoholischen Lösungen, und auch in der Netto-Menge des konsumierten Alkohols. Des Weiteren wurde eine Entwicklung des Trinkverhaltens im Laufe der aufeinanderfolgenden Zyklen von Alkoholkonsum und Entzug beobachtet, mit einem klaren "Einsteiger"-Profil und der Entwicklung zu einem mehrerer fortgeschrittenen Profile. Mit Hilfe des hier vorgestellten Ansatzes können Unterschiede im Verhalten für verschiedene Bedingungen und/oder Gruppen von Tieren identifiziert werden, und sogar eine Vorhersage von Alkoholismus bereits in frühen Phasen des Alkoholkonsums ist möglich. Die entwickelten Methoden sind auch in zahlreichen anderen Gebieten anwendbar, in denen die Analyse von Zeitreihen eine wichtige Rolle spielt (u.a. Analyse von Microarray-Daten, Neurowissenschaften, etc.).

Contents

Chapter 1

Introduction

Alcoholism is one of the most critical drug addictions of our society, carrying with it not only several social and health problems but leading, in many cases, to crime and death.

The development of alcoholism, as most of the drug dependencies, occurs without notice. Only once some acute features have appeared, such as the compulsive seeking and consumption of alcohol leading to intoxication, craving and relapse during withdrawal, amongst other, individuals become aware of the possibility of being addicted. By this point, complete recovery from this illness is unlikely. Understanding all the social and biological aspects of this disease is important, in particular, so that, from an early stage, we can describe and predict the risk of developing it.

Many different directions have been taken, with the main objective of gaining insight on alcoholism. Studies on humans have brought to light the relationship between alcoholism and genetic and environmental factors [JvdBGP98]. Those however are limited in many senses regarding human testing ethics (e.g. drug testing) or are costly and affected by large variation due to self reports (e.g. followed studies based on self-reports, to learn how an individual develops into an addiction).

Animals can be used to obtain a deeper knowledge of the processes underlying alcoholism. In-cage studies allow the extraction of patterns of behavior towards alcohol intake, how they vary in time, how they are affected when other drugs are administered, etc. This also has, however, limitations, since drug dependencies are strongly influenced by social aspects, and are therefore difficult to simulate under laboratory conditions.

There is a large body of research on the behavioral study of alcoholism in animals including (among others) how it develops, which physical symptoms it provoques and signs of withdrawal due to abstinence.

The first research on *animal models of alcoholism* aiming to induce and maintain a dependency on alcohol in animals, dates back to the late 1960s [Mel76] and has kept evolving until today. In 1973, Lester and Freed summarized a set of features that such an experimental protocol should expose in order to be called an animal model of alcoholism. Refined scenarios have

been designed based on the guideline given in their paper [LF73]. Main goals are to induce in animals sustained high alcohol intake and blood levels, as well as, withdrawal related symptoms.

A documented feature of withdrawal due to abstinence from alcohol of humans and many animals (including rodents and monkeys) is the so called *alcohol deprivation effect (ADE)*. It is related to an increased ethanol intake after a period of abstinence. In the case of rodents, this phenomenon has been widely studied [CMB68, SSJ73, Eri72, Gol72b, Gol72a, GA76, HSW+02] and mathematically modelled [SS67]. If quinine[1] taste adulteration of the alcoholic solutions after a deprivation period does not affect the presence of an ADE (in spite of its aversive taste), then it can be considered to be a symptom of compulsion and loss of control in the alcohol drinking, thus being referred as a sign of an alcoholic dependence in rats.

The *long-term alcohol self administration with repeated alcohol deprivation phases* [SH99] is an animal model of alcoholism aiming to simulate in rats the process of free alcohol consumption and withdrawal after a deprivation phase. Rats receive for long periods (4-5 weeks) free choice of water and ethanol $5\%, 10\%$ and 20% concentrated solutions. They are then deprived for a week from the alcoholic solutions and afterwards get the alcohol represented. After several such baseline-deprivation-representation phases, rats are supposed to present ADE even in the presence of quinine adulterated solutions.

The use of a novel *drinkometer* device, allowing the recording of a high definition drinking time series, provides this thesis with valuable data sets. So far, the analysis of this data has been limited to descriptive observations such as mean ethanol intake and its variation after a deprivation phase (e.g. solution preference development throughout time in blocks of several weeks, impact of drugs on ethanol increase related to baseline, after-deprivation [SHA+96, SH99]).

In this thesis we extend the approach of [SHA+96, SH99] to include both the identification of particular drinking patterns within a population, and provide a high temporal resolution description of their dynamics. We are furthermore interested in studying the evolution of the regular drinking patterns and withdrawal symptoms throughout time. This approach provides this thesis with interesting time series classification and modelling problems.

Several time series clustering approaches exist. [WL05] categorizes them into raw-data-, feature- and model-based-clustering. In this thesis we focus on the latter, the most advanced developments of which are reviewed in [FS11]. Under this framework, a population is represented by H sub-populations, each of them described by a parametric model. This can be mathematically defined through the concept of *finite mixture of distributions* [MP00].

The fitting of generic mixtures of distributions is achieved through the general purpose framework proposed by Dempster et al. [DLR77] for maximum likelihood estimates with hidden data. They propose the *estimation maximization* (EM), which can solve, as one of its many applica-

---

[1]rats find quinine taste adulterated ethanol solutions aversive, because of its bitter taste, causing a decrease on the intake

tions, finite mixtures. EM has several advantages and disadvantages. The advantage is, it is simple and intuitive to develop and requires little time to give a solution. The disadvantage is that it converges to a local maximum, not yielding optimal solutions only by good starting approximations (or, as proposed here, restarting several times and keeping the best solution).

Through EM, mixtures of distributions can be fitted. The only remaining question is how to model sub-population dynamics. For this, *generalized linear models* (GLM) are chosen, since they can represent a wide range of known distributions, i.e. the *family of exponential distributions* (EM) [Cla05], and are intuitively generalized to time series. In [WD95], a procedure for the clustering of multivariate data based on a GLM-EM procedure is proposed. We generalize their results for the case of multivariate time series, yielding a procedure for simultaneously classifying time series into several clusters and modelling the dynamic mean behavior of each observed group in terms of a GLM.

Summarizing, we propose a framework to model drinking behaviors. A behavior is represented by the probability of a drinking event a given time point of the day. It resembles a Fourier series (it is a non-linear function of a linear combination of sines and cosines). The best selected model (through CAIC [WD95]) has periodic components reflecting day/night cycles and other higher frequency components (most likely reflecting different stages of the alcohol metabolism). To complete the approach, several such behaviors are assumed in the population. With the aid of an EM, a mixture of the described behaviors is fitted to the data, resulting in a partition of the data set, whose analysis allows us to estate that rats drink alcohol in several ways, which evolve in time from explorative to advanced solution preference profiles. Relapse behavior is also modelled under this framework, showing how it tends to last longer and become inflexible[2] throughout time.

1.1 Hypotheses

Since the data sets analyzed for this thesis come from Wistar rats, the hypothesis are related to these animals. However, they could be extended when new data sets from different animal types become available.

1. Wistar rats have determined patterns of behavior towards alcohol which develop over time.

2. After a first deprivation phase, Wistar rats will tend to drink ethanol intensively on the first day after representation of alcohol. However, if the *EtOH* solutions taste is affected with bitter substances, particularly with quinine, they will not drink intensively at this point.

3. After several deprivation phases, rats develop inflexible drinking in presence of quinine-adulterated alcoholic solutions, and thus indicating a dependence on alcohol.

---

[2]in spite of aversive taste of alcoholic solutions, rats drink as much as if the solution were not adulterated

4. Alcohol deprivation effect, a sign of alcoholic addiction, characterized by loss of control and compulsive alcohol intake after a deprivation phase, can be predicted from early intake stages.

1.2   Summary of results and contributions of this thesis

A group of 29 male Wistar rats are studied, that were drinking in-cage under the long term self administration with repeated deprivation phases protocol on a four-bottle paradigm (see Chapter 3). High resolution time series were recorded using the *drinkometer* system, during different baseline/after-deprivation phases.

Our first contribution involves the proposal of an H_2O - *penalized alcoholic intake* measure for the four-bottle paradigm. Several ethanol intake measures found in the literature are discussed and the new one is proposed, overcoming most of the observed disadvantages.

Based on the proposed intake measure, a procedure to classify an animal into presenting ADE or not is described. It is based on comparing the increase in intake after a deprivation phase of controls and quinine treated animals. Such a classification does not exist in the literature, since in most cases mean intakes are compared between groups (e.g. preferring/non preferring, control/treatment), but a single animal is not classified into presenting ADE or not.

We develop a methodology to simultaneously classify and characterize dynamic patterns of drinking behavior of time series recorded under a free choice of water and several alcoholic solutions paradigm. This is done by

- extending known results on *generalized linear models (GLM)* for univariate time series to the multivariate case, and simplifying the computational fitting procedure assuming a shared seasonal pattern throughout individuals (see Chapter 4) and

- implementing an *estimation maximization* algorithm to fit mixtures of the mentioned multivariate GLM (see Chapter 5).

The result is a partition of the data as well as a characterization of each group in terms of a GLM.

The designed procedure is used to analyze the described data set. Patterns of behavior are found in the first baseline that can be described as naive/primary, since no particular preference for an alcoholic solution and a high frequency of drinking events (mostly of water) can be observed. In further phases, advanced patterns showing a preference for a solution and a decreased frequency of overall drinking events characterize most of the rats. Those patterns were observed for different rats also at advanced baseline phases, proving the robustness of the proposed methodology.

It can be seen how primary patterns (observed mostly during the first baseline phase) will result in a variation in the ADE, conditioned on the presence of quinine. I.e. controls will present ADE while the quinine group will not. On further phases, regardless of quinine, most of the animals drinking under an advanced pattern will present ADE. So already from a baseline phase

the ADE outcome can be predicted.

However, the homogeneity of the Wistar rats in presenting ADE at advanced drinking phases does not allow to explore beyond in the risks of presenting an alcoholic addiction from early patterns of intake. It can be already predicted, before performing any behavioral experiment, that after several deprivation phases, Wistar rats develop an inflexible alcohol drinking pattern, thus presenting ADE regardless of the bitter taste of the alcoholic solutions.

If a similar data set were provided, where the ADE is not guaranteed after several deprivation phases for all the population, but more a random effect (meaning that some rats will end up alcoholic and the rest not), the whole procedure could be employed without much further effort. A correlation analysis could be performed to establish whether determined patterns of behavior at early drinking stages condition the development of the addiction.

With this we would like to stress the fact that the proposed methodology is generally applicable and so can be employed beyond the scope of the data considered in this thesis. It allows one to predict future outcomes given early behavior patterns. For this, we would require:

- time series data set describing an early phase of the phenomena of interest and

- a classification of the advanced state (e.g. alcoholic/non-alcoholic) for each individual

Our methodology provides a way of partitioning the individuals into a few groups, each of them dynamically characterized by a model. The identified groups can be correlated to the given classification. Conclusions about the advanced states can be expressed as probabilistic distributions dependent on the observed patterns.

1.3 Structure of the thesis

This thesis is structured in 7 chapters and 2 appendices.

Chapter 2 provides a wide literature review on the biological as well as methodological fields, indicating where this thesis is placed in the current state of the art.

Chapter 3 introduces the application that motivates this thesis. It describes the *long-term alcohol self administration with repeated deprivation phases* protocol under which the data is recorded, the *alcohol deprivation effect (ADE)*, the recorded time series as well as its preprocessing procedure. The chapter finally proposes:

- a new measure of the alcohol intake under the 4-bottles free choice paradigm, and based on it,

- a method to determine, whether rats present ADE or not.

Chapters 4 and 5 are devoted to the description of the statistical modelling and clustering framework. Appendix A comprises a simulation study that asserts the accuracy and performance of

the proposed framework. Model fitting and parameter selection as well as clustering performance are tested using simulations, yielding very good results given that distributional assumptions hold.

Chapter 6 discusses and interprets the inferred patterns of behavior. An evolution of the drinking patterns can be observed throughout time, as well as an acuteness of the loss of control symptoms in the alcohol drinking, following a phase of abstinence. Robustness of found patterns is tested by introducing a new data set obtained from an advanced drinking phase under the same laboratory setup conditions. Since similar behaviors are observed, we conclude that the observed patterns are robust. Further figures that might help in the interpretation (particularly for the biologists) are presented in Appendix B.

Chapter 7 summarizes the obtained results and sketches future work in both data and methodology directions.

Chapter 2

Scope and state of the art

2.1 Animal models of alcoholism

Only humans voluntarily drink alcohol to intoxication, and continue to regularly use it, in spite of the negative social and health implications. A major goal of alcoholism research is to understand the mechanisms that cause humans to undergo such a self-destructive process.

Several studies are considered unethical when the subjects of the experiments are humans. The use of animals to obtain a deeper knowledge of the underlying process of the dependency gives scientists a strong tool for the research. Examples of the subjects of experiments in this field in the research on alcoholism are primates [Sin71, RL74], mice [Gol72b, Gol72a, GA76, CPF$^+$96] and rats [Ric26, Mye62, SS67, CSPS71, Eri72, CS73, CAL$^+$95, SHA$^+$96, SH99].

A limitation in the use of animals for alcoholism research, is that they do not freely drink alcohol. They are less prone to intoxication due to excessive ethanol consumption and to developing an addiction. Animals must be induced to consume alcohol. Furthermore, alcoholism-related features have to be identified in order to be able to state that an animal has a dependency. Such features include tolerance (need more alcohol to still feel the effect), psychological dependence (have a desire for alcohol) and withdrawal symptoms (physical reactions during abstinence). These represent some of the elements of the classical criterion of addiction to alcohol.

In the following we present a literature review, aiming to summarize some of the most remarkable historic developments in the use of animals for alcoholism research.

The first developments on alcoholism in animals date back to 1919. At the time it was shown that rats can develop an acute tolerance to alcohol, and are able to perform tasks after high doses of alcohol intake [Mel19]. These results turned out to be extremely important, implying that rats can present signs of physical dependence to alcoholism (such as tolerance).

Humans consume voluntarily alcohol and so develop into an addiction. Behavioral studies of this illness (patterns of intake and their changes conditioned on external stimuli) on animals, require that the subjects follow a similar pathway towards the consumption of alcohol, as well

as, that they present symptoms of physical dependence. This was early recognized, so much attention was drawn towards the different factors inducing a self-selection of alcohol.

Some preliminary results on this topic showed that white rats drank voluntarily alcohol when an alcoholic solution and water were simultaneously offered (given that they had undergone an initiation procedure involving the presentation of alcoholic solution and no water at a very early age (25-30 days)) [Ric26].

In 1960, the changes in the preference for alcohol induced by several experimental set-ups like different types of rats, diet taste and smell variation, introduction of a third choice (e.g. sugar, fat) and drug influence are reviewed [Mar60]. It is a first characterization of the different experimental settings of alcoholism in animals, serving as a base for further advances in the topic. Further characterizations showed how temperature and ethanol concentration influenced the preference for alcohol in G-4 and Wistars rats: more alcohol was drunk at $27°C$ than $18°C$; rats avoided high concentrations of alcohol (20%), and in the range $1.25% - 5.0%$ the intake of alcohol was in general greater than water [Mye62].

In 1968, a relationship between a psychological stressor and the conscious choice of ethanol is established: rats were receiving clued and non-clued shocks and the ethanol intake was measured. An increased ethanol intake was observed when rats were warned in advance of the occurrence of a shock, while the non-clued shock did not provoke significant increase [CMB68]. This had a great impact, implying that rats consume alcohol when they are stressed, i.e. for its pharmacological effects.

Although the mentioned approaches were considered to be signs of alcoholism in animals, all of them seemed to fail in inducing important features of alcoholism, such as maintained ethanol intake or withdrawal symptoms. In 1971, the first model for alcohol addiction in the rat which satisfied the classical pharmacologic criteria of addition (physical dependence, tolerance and indifference) was proposed [CSPS71]. Here, twenty-four rats of the Holtzman strain were given either water or a 7 per cent alcoholic solution from 21 days until 154 days of age, and food ad lib. They were afterwards deprived from alcohol showing highly anxious behaviors in contrast to water rats. Alcohol rats presented a tolerance to alcohol with respect to controls, when injections of alcohol were provided to both groups. The intake of alcohol increased during the first exposure period and seemed to remain constant, when competing solutions were offered.

In 1973, Lester and Freed, after reviewing the approaches to developing alcoholism in animals, gave a list of criteria (we reproduce them in Table 2.1) to be fulfilled in the design of an animal model of alcoholism [LF73]. This paper is still a guideline for all scientists working in the field.

Though Table 2.1 gives a list of desired features of an animal models of alcoholism, there are many aspects that can not be simultaneously considered in experimental set-ups [Poh81]. This is why different experimental models focus on a subset of those features while weakening the others. The advances in the field until 1976 are classified into pharmacological (forced alco-

1. Oral ingestion of alcohol without food deprivation.
2. Substantial ingestion of alcohol with competing fluids available.
3. Ingestion directed to the central intoxicating character of alcohol, substantiated by determination of circulating blood alcohol levels
4. Work performed, even in the face of aversive consequences, to obtain alcohol.
5. Intoxication sustained over a long period.
6. Production of a withdrawal syndrome and physical dependence.
7. After abstinence, reacquisition of drinking to intoxication and reproducibility of the alcoholic processes.

Table 2.1: Criteria for an animal model of alcoholism [LF73]

hol administration) and behavioral (alcohol self-adminstration) models of alcoholism [Mel76]. Amongst the applications of both types of experimental branches are the first model of an alcohol-induced cirrhosis in baboons [RL74], and the relations between withdrawal syndrome intensity and duration, and the alcohol dosage and exposure-duration [Gol72b, Gol72a].

The recent results in the field of *animal models of alcoholism* were reviewed in [Spa00]. Three main directions were described:

- The *alcohol preference models*: several alcohol preferring/non preferring animal lines have been developed through selective breeding, and a whole characterization of their behavior towards alcohol is widely documented. That is the case of the AA/ANA lines of rats [Eri72], the Sardinian alcohol preferring rat [CAL+95], the HAB (high anxiety-related behavior) and the non-anxious LAB (low anxiety-related behavior) lines [HSW+02], and the alcohol preferring p rats [BRL+06]. These lines allow the correlation of genetic factors with alcohol drinking behaviors [CdPBDF05, CPF+96]. The fact, however, that a high intake of alcohol does not by itself indicate the presence of an addiction is stated. In [ML98], several approaches to alcoholism in rodents are reviewed, which compare important features between pairs of high vs. low alcohol preferring rodent strains (e.g. self ethanol intake, withdrawal syndrome, tolerance, etc).

- The *reinstatement model*, where animals are trained to receive a drug in response to a task. Then the drug is withheld, even if the animal keeps performing the task. After this, the animal ceases to perform the task. It has been proven that for rats, stress, injections of a small dose of the drug and conditioned stimuli paired to the training session reinstates the seeking for the drug [LQJ+98, KMW99].

- The *long term alcohol self-administration with repeated alcohol deprivation phases* is the third proposed model. This one will be described in Chapter 3 of the thesis and its origins will be described in the next section.

While genetically selected animals are valuable in the identification of significant genes conditioning an addiction, the use of "wild" animals (though still inbred for laboratory purposes) to uncover the natural pathway to developing an addiction is of interest. The latter represent

a closer model of dependency in humans, which can develop alcoholism independently of their genetics. This is why many behavioral studies are based on this type of animals.

2.1.1   Alcohol deprivation effect

A documented feature of withdrawal due to abstinence from alcohol of humans [BMCS81] and many animals (monkeys [Sin71], mice [SS93]) is the so called *alcohol deprivation effect (ADE)*. It is related to an increased ethanol intake after a period of abstinence. In the case of wild rats, this phenomenon has been widely studied by Sinclair et al. [SS67, SSJ73]. The increase in ethanol intake after a deprivation phase with respect to baseline drinking is modelled in terms of an exponential decay on the days after representation of alcohol. The effect of long and short alcohol deprivations in two strains of alcohol preferring rats is analyzed in [SL89]: the AA rats [Eri72] did not present any sign of ADE after a week deprivation, however, after only one hour of deprivation, their intake increased significantly with respect to controls. P rats [BRL$^+$06], on the contrary, presented a similar ADE as control Wistar rats [Kin18] after a week of deprivation and a much higher increase during the first hour of representation, following few hours of alcohol deprivation (like the AA rats).

The *long-term alcohol self administration with repeated alcohol deprivation phases* [SH99] is an animal model of alcoholism to simulate in rats the process of free alcohol consumption and withdrawal after a deprivation phase. Rats receive for long periods (4-5 weeks) a free choice of water and differently concentrated alcoholic solutions (e.g. 5%, 10% and 20%). They are then deprived for a week from the alcoholic solutions and afterwards get the alcohol represented. After several such baseline-deprivation-representation phases, rats are supposed to present ADE even in the presence of quinine adulterated solutions or a competing palatable solution [SHA$^+$96]; this shows the potential of this model to simulate important symptoms of an alcoholic addiction: compulsion, loss of control and indifference in the alcohol drinking.

2.2   Alcohol drinking patterns and mathematics

Mathematical tools have been used by biologists to assert their intuitive results. For example, descriptive statistics (e.g. mean, variance and correlation) have been used to gain an insight of the overall phenomenon, and statistical tests have been used to assert, to a certain level of statistical significance, the observed features.

In the field of alcoholism, the term drinking pattern has been used in a very ambiguous way, mostly related to descriptive statistics on intake, inter-drinking intervals and/or drinking frequencies [GA76, MLE$^+$76, BRS$^+$06].

However, in 1994, Grünewald and Nephew [GN94] provided a more mathematical approach by defining drinking behavior applied to a data set obtained from a general population study of California consumers. They modelled the probability of a drinking event in terms of a logistic

function and showed that the frequency of drinking distributed exponentially. [Gru98] reviews the tools for modelling the distribution and consequences of alcohol consumption. Here, the *stochastic drinking theory in the modelling of drinking behaviors and drinking risks* is defined. The basic goals of this approach are:

1. to provide a universal description of individual drinking patterns in probabilistic terms

2. to provide the linking theory to relate the patterns with alcohol-related harmful outcomes

3. to provide a theoretical explanation for the origination of drinking patterns from an ecological constraint.

[GRL$^+$02] introduced more recent advances in mathematical modelling of current drinking patterns, drinking disorders and their evolution throughout life in humans. The results discussed were mostly based on self-reports of daily drinking patterns, thus highly affected by external factors. Though results are given for humans, they can be extrapolated to animal research, where the information on the drinking quantities, frequencies, etc. are less affected by large variations. Moreover, in-cage studies of animals can provide reliable (unaffected by external environment) high time resolution data, allowing a deeper insight in the drinking patterns.

The use of a novel *drinkometer* device in the laboratories of the ZI-Mannheim provides this thesis with panels of high resolution multivariate time series from Wistar rats drinking alcohol. Following the same goals as Grünewald, we focus on finding drinking patterns within a population, analyzing their evolution and relating them to alcoholism-related outcomes. This framework will provide scientist in this field with a tool to recognize how early evolutions in the drinking patterns can condition further risk of presenting alcoholism. To our knowledge no developments of this kind have been described in the available literature.

2.3   Time series - generalized linear models

Among the first developments in the analysis of time series data, was performed by Beveridge [Bev21], who studied how weather conditions affected wheat price fluctuations in Western and Central Europe. Here, harmonic decomposition is applied to data that was measured from 1500 to 1869. It was concluded that *"the yield of harvests in Western and Central Europe from the middle of the sixteenth to the opening of the twentieth century has been subject to a periodic influence or combination of such influences tending to produce bad harvests at intervals of about 15*3 years, the first epoch falling in 1556"*. This allowed furthermore to predict that the harvest of 1923 should be deficient. The validation of this statement is complicated by political factors, as the first world war also influenced the wheat market. Indeed very depressed prices were given in the whole first half of the 1920s [BG03].

Developments in time series modelling can be classified in several ways (namely by time [BGJ73, Bol86] vs frequency [Blo76, PW00] domain, normal vs non-normal distributed residuals, stationary vs non-stationary, etc.). It is difficult to make an exhaustive historical review, since this

area of research has developed parallel in different research fields. We therefore jump directly in the methods for either multivariate or categorical time series, that finally led us to choose generalized linear models for our modelling purposes.

First developments on categorical time series modelling are related to Markov Chains [Tau86], integer autoregressive models [JGY91] and discrete autoregressive moving average (ARMA) models [PP81].

The *generalized linear model* (GLM) framework was introduced by Nelder and Wedderburn [NW72] in 1972, and considers random variables following a member of the *family of exponential distributions* (FE) [And70]. It provides a unifying framework for the modelling of random responses linked with systematic effects without necessarily assuming normally distributed residuals. For a detailed definition of the model, parameter estimation algorithm and appropriate starting guesses, as well as several generalizations and example applications see [MN89].

The FE contains in its definition many of the most commonly used distributions. Although first defined around 1935, it is convenient for the development of GLM, an has therefore attracted much attention. A good review on its definition, its moments (mean, variance and higher order ones), and some example members (e.g. continuos distributions, such as the normal and exponential, and discrete distributions as the binary, poisson and multinomial) is provided in [Cla05].

Special uses of the GLM in the autoregressive modelling of categorical time series have been developed by Fahrmeir and Kaufmann [FK87] (extending regression models for stochastically independent observations to allow for non-stationarity), Pruscha [Pru93] (introducing covariates via logistic regression) and Fokianos and Kedem [FK98] (considering categories as states and liking their probability of occurrence to the covariates through a time-invariant parameter vector).

A generalization of GLM for expressing the existence of several latent classes is presented by Wedel and DeSarbo [WD95] with the help of the definition of mixed models and estimation maximization (EM). They show furthermore how several of the latent-class regression methods can be seen as special cases of their proposed models. A study is developed and conclusions are drawn regarding goodness of fit and model selection as well as a wide range of example applications.

GLM is important for multivariate analysis, since several of the known multivariate distributions can be also considered a member of the family of exponentials (e.g. multinomial, multivariate normal). The book of Fahrmeir and Tutz [FT01] provides a useful introduction with many example applications, covering from the simplest definition of GLM for cross and long sectional studies, univariate and multivariate responses, time series and mixed models.

2.4   Time series model-based clustering

Several time series clustering approaches exist. [WL05] categorized them into a) raw-data-, b) feature- and c) model-based clustering. Since this thesis focuses on the latter, we will discuss it in detail below.

Model-based clustering focusses on assuming that (1-) a population can be partitioned into sub-populations and (2-) each of them can be described by a determined model. The concept of *mixture models* marries both these requirements, thus being the basis for model-based time series clustering methods.

The idea of *mixture of distributions* was already employed at the end of the 19th century in [New86] and [Pea94]. The latter proposed a mixture of two univariate normals to describe some data from crabs, fitting the model through a moment based estimator, yielding complex calculations he performed without the aid of a computer. In 1920, Green and Yules discussed the "frequency distributions representative of multiple happenings", questioning if diverse outcomes of a disease should be seen as a single probabilistic process, or as the result of the effects of different environments yielding different such models [GY20]. On 1941, the papers from Halmos et al. [Hal41, AHK42] developed theoretical results for the decomposition of a measure space into the sum of measure spaces. These formed the basis for the later work of Robins [Rob48], where he formalized the concept of *mixture of distributions*, as a way of representing several subgroups within a population.

There have been many attempts at solving the problem of efficiently fitting a mixture of distributions. These were typically done for very specific cases. As an example, [Has69] proposed a maximum likelihood estimator for mixtures of distributions coming from the exponential family, and [CM73] proposed one for mixtures of discrete probability functions.

In [DLR77], an *Expectation Maximization* algorithm for incomplete data is formalized and several applications are described, amongst them the fitting of mixtures models. Its simplicity, adaptability, and general good performance makes it one of the most commonly used inference techniques in presence of incomplete data. Particular examples are missing data [Har58, CBM71], finite mixtures models [Has69, CM73] and hyper-parameter estimation.

A complete description of mixture models, including applications and a historic review is given in the book by McLachlan and Peel [MP00].

As pointed out before, finite mixtures has as one of its main applications the model-based clustering. Particularly in the field of time series, it has had a key role in the last decade.

In [FS11], the most modern time series model based-clustering techniques are reviewed. All of them have the common need of defining a modelling kernel to characterize each subpopulation and a prior distribution for the class assignment, giving place to the different discussed approaches.

Given the type of the analyzed data, a complete branch of model-based classification has been developed. Real-valued time series are mostly represented by dynamic autoregressive models and have been applied to financial [FSK06] and population growth problems [OVD09]. Binary time series kernels are usually based on logit [ZZ04] or probit [ABH11] models, describing probabilities of success at time t. Count time series are generally modelled through a Poisson distributions, and general discrete time series by terms of Markov Chains [CHM+03, RSC02, PS10]; information on covariates lead to a multinomial log-it clustering kernel (for modelling categorical but not multivariate data) which has been applied on [WD03].

Mixtures of GLM for the partitioning and modelling of multivariate data (although not time series) are introduced in [WD95]. They make the assumption of independence between responses to simplify the fitting process. The mentioned model has been used on functional neuro-imaging data [PF03] to obtain clusters of voxels characterized by similar time series, allowing the discovery of task-related connectivity between different parts of the brain.

Motivated by the above mentioned methods from [FSK08, FS11] and [WD95], in this thesis we develop an Expectation Maximization algorithm, which is adapted for the simultaneous clustering and fitting of mixtures of GLM for **multivariate long-term time series**.

Chapter 3

Animal drinking time series and ADE

Alcoholism research based in animals has many difficulties:

- Since animals mostly do not drink voluntarily alcohol (not even humans), how can they be induced to do it to such an extend that they will become alcoholics?

- Once they do: how can researchers know that the animal has developed an alcoholic dependency? Even to assert that a person is alcoholic, several questionnaires have to be filled in and several social disorders have to be observed.

With this motivation, the so called *animal models of alcoholism* have been developed, where the subjects proceed through certain scenarios intended to induce alcoholism (e.g. stress factors, deprivation from food). In presence of an addiction, they then react similarly to addicted persons in homologous situations. [LF73] describes seven criteria to be fulfilled in the design of such models, s.a. the voluntary consumption of ethanol without food deprivation, the presence of physical dependence and abstinence symptoms during withdrawal.

[Mel76] presented a review on the first developments of this area of research. It characterizes the animal models of alcoholism into two groups:

- pharmacological: forced alcoholic administration to induce a physical dependence. This approach allows a very fast development into an addiction and is useful in dependence-related studies, like drug testing, abstinence symptoms on withdrawal, etc. It is, however, not appropriate for studying the process of developing the addiction per se.

- behavioral: self administration either free or through a task to obtain the alcohol. Examples of such models are the *oral ingestion through schedule-induced polydipsia*[1] and *the intravenous self-administration.*

At the time, the reviewed behavioral models seemed to need a stimulus in order to induce alcohol intake, and mostly lacked maintained alcohol intake levels when the stimulus was eliminated.

---

[1]polydipsia: intensive thirst

A more recent review on advances in this topic is given by [Spa00]. Some of the described protocols are: the *reinstatement model* as a measure of craving and relapse in the alcohol drinking; the *point of no return model* through which it has been proven that a point exists where rats start drinking in an uncontrolled way, and even after a long period of abstinence, they will remember and relapse into alcohol drinking; and the *long-term alcohol self administration with repeated alcohol deprivation phases*. The thesis is based on the latter, which is described in more detail in the next section.

3.1 Long-term alcohol self administration with repeated deprivation phases

The *long-term alcohol self administration with repeated alcohol deprivation phases* [SH99], is designed to observe relapse-related features following a withdrawal phase, amongst which one can find: enhanced drinking, loss of control and anxiety.

Under this protocol, in-cage Wistar rats are offered four bottles (water and alcoholic solution at concentrations 5%, 10% and 20%[2]). They have free access to all of them for 4 to 5 weeks. The drinking amount is measured every 5 minutes with the help of an automated device (*drinkometer*) yielding long time series (20 days of 288 daily measurements). This represents a *baseline drinking phase* (B). After this time, the rats are deprived from alcohol for 2 weeks, during which they only receive water and food. This is called the *alcohol deprivation phase*. Then the alcohol is re-presented and measurements are taken again for a week. This is called the *after-deprivation phase* (AD). Figure 3.1 shows graphically the described protocol. This procedure is repeated several times yielding subsequent baseline, deprivation and after-deprivation phases in the drinking lifetime of the rat.

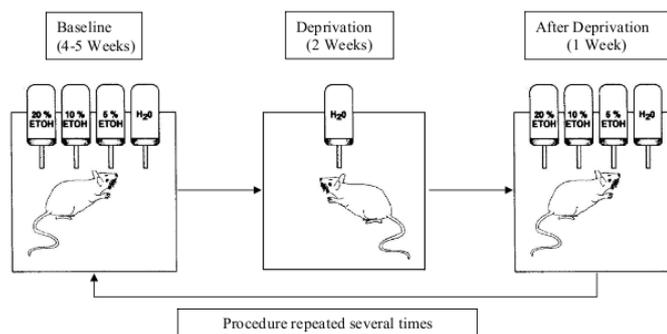Figure 3.1: Long-term alcohol self administration with repeated alcohol deprivation phases

During the AD phase, rats are expected to present the so called *alcohol deprivation effect* (ADE), characterized by a change in the pattern of consumption and thus an increase in the mean al-

---

[2]ethanol solutions will be notated in the following as *EtOH*, followed by the percentage concentrations, e.g. *EtOH*5%

cohol intake with respect to the baseline for the first few days of alcohol re-presentation. Using specific tests, like alcohol taste adulteration, or presentation of other tasty solutions, it can be shown that rats indeed drink without control and seek for the effects of the drug. This is one of the most desired features of an animal model of alcoholism, since it relates to the alcohol dependence symptoms of humans.

A characterization of drinking under this protocol has already been undertaken: in [SH99] our collaboration partners carried out a study on male Wistar rats where:

- a comparison between the described 4 bottle and a 2 bottle (water and $EtOH10\%$) variants was made, where the former achieved the highest ethanol intake and preference;

- changes in alcohol drinking patterns over time were observed in terms of total amount of consumed alcohol and preference to the alcoholic solutions. It could be seen how the initiation phase was characterized by the highest alcohol consumption (in g of alcohol per body kg), and throughout time decreased and finally stabilized. The general preference for ethanol remained constant, though an analysis on the single solution showed a variation (at the beginning $EtOH5\%$, $EtOH10\%$ and $EtOH20\%$ had a mean preference of 30%, 15% and 10%, respectively, while at the end of the study the preference evolved to approximately 14%, 19% and 19%, respectivel;

- the presence of physical and psychological signs of withdrawal could be observed during a deprivation phase, e.g. hyperlocomotion and hypothermia, increase in anxiety-like behavior measured through the social interaction and elevated plus maze tests. Most of the observed symptoms were reported to become more acute in long-term drinking rats (after several phases of alcohol deprivation).

However, the description of different patterns of drinking behavior, their evolution in time, how the regular way of drinking induces a relapse pattern after abstinence, amongst many others, has not yet been performed. The use of mathematical modelling tools will contribute to achieve the following goals:

- Identification and characterization of patterns of behavior during measured baseline and after-deprivation phases.

- Characterization of a possible long-time evolution in the drinking patterns.

- Correlation of the found patterns of behavior during baseline and after deprivation phases to fully characterize and later predict the ADE as a sign of alcoholism.

This chapter refers to the description of the measured data throughout phases; the *alcohol deprivation effect* (ADE), considered under determined circumstances a sign of alcoholic dependence; the classification of each animal as presenting or not ADE. As a part of this chapter we propose a way to measure the intake under the four-bottle paradigm, based on a H_2O penalized net $EtOH$ intake.

ADE is a commonly observed behavior in many animals after a period of alcohol abstinence. It is characterized by an increased ethanol intake, and increased locomotion and anxiety levels during the first days of representation of the alcohol.

In the case of rats, the increased intake after-deprivation with respect to baseline drinking levels was mathematically modelled by Sinclair et al. [SS67, SSJ73]. The simple model states the exponential drop of the increase, either of the net $EtOH$ intake or the $EtOH$ preference, as function of the amount of days after representation of alcohol, i.e.

$$AD_Increase(d) = M(e^{-a*d}) \tag{3.1}$$

where $d \in \{0, \ldots\}$ is the amount of days after representation of alcohol and M and a are the model parameters to be fitted to the data. Parameter values are given that are reported to account for 98% of the daily variation. Those studies were performed under the two-bottles choice protocol (H_2O and $EtOHp\%$, being $p = 10$ commonly used) and based on home cage measurements.

The inclusion of the four-bottles paradigm as well as the novel device for high time resolution measurement of solution intake, performed in the labs of ZI-Mannheim, has uncovered some variation in the behavior of Wistar rats towards a deprivation of alcohol. The analysis itself has to be adapted, since the usual net ethanol intake (g/kg) or solution preference can not meaningfully explain all the information given. Details are presented in Section 3.2.

Since our later and final goal will be to predict alcoholism from early intake phases, a classification of an animal into addicted or not is necessary. Quinine taste adulteration of the alcoholic solutions is a technique to assert compulsion and loss of control on ethanol drinking, both of them signs of an alcoholic dependence. Though much has been written on ADE, very little has been said about a classification of an animal into presenting ADE in presence of quinine. Section 3.3 addresses this matter, based on the fact that an animal drinking after-deprivation from quinine-alcoholic solutions at the same basis of controls should be classified as presenting ADE. In spite of the aversive taste, they increase the baseline mean intake at the same rate as control animals, showing a loss of control and compulsion in the drinking, thus, an addiction.

3.1.1   Data description

The data analyzed in this chapter and in Chapter 6 was recorded under the described protocol. Drinking time series of 30 male Wistar rats during the 1st, 3rd and 5th baseline and after-deprivation phases were measured. Each time series consists at every time point of a 4-valued real vector, containing the amount of solution a rat consumed from each bottle during the previous 5 minutes. Figure 3.2 shows a portrait of the measured data.

During the 1st AD, half of the rats received quinine in the alcoholic solutions. During the third phase, no animal received quinine. On the 5th AD, the other half got the quinine adulterated

alcoholic solutions.
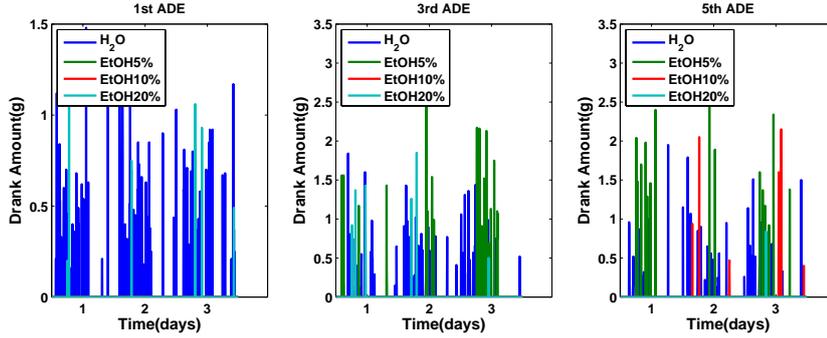
Figure 3.2: Portrait of measured time series

In order to validate the robustness of the obtained results, as well as observe how advanced animals drink during the fifth phase compared to more advanced phases, data from 22 different rats on the 9th cycle is added to the analysis. This data set is also used in the ADE classification procedure.

3.2 Alcohol intake measure

The two-bottles free choice paradigm (one bottle of H_2O and one of $EtOH$, usually 7% or 10%) has been used in several environments. [SS67, SSJ73] even model ADE based on measures developed for this situation. However, in the ZI-labs, a four-bottles paradigm is used. In [SH99] the advantages of this protocol in terms on an increased ethanol intake and preference in the initiation phase with respect to the two-bottles are described. The existing measures have to be extended to allow the compilation of all the given information.

In the following, several approaches to solve this issue are given, together with their advantages and disadvantages. We call them *alcohol intake measure*. Figure 3.3 gives a graphical overview of each of them, and Figure 3.4 shows the mean daily intake under the two most meaningful representations throughout successive baseline phases.

3.2.1 Alcohol preference

- Given the amount of drank H_2O and each solution $EtOH_p$ $p = \{.05, .1, .2\}$, the preference is calculated as

$$Pref(H_2O) = \frac{Am(H_2O)}{Am(H_2O) + \sum_{p \in \{.05,.1,.2\}} Am(EtOH_{p\%})} \qquad (3.2)$$

and

$$Pref(EtOH_{p\%}) = \frac{EtOH_{p\%}}{Am(H_2O) + \sum_{p' \in \{.05,.1,.2\}} Am(EtOH_{p'\%})} \qquad (3.3)$$

- *Equations (3.2) and (3.3) show how animals prefer a solution but does not show the fraction of alcohol intake each solution provided.*
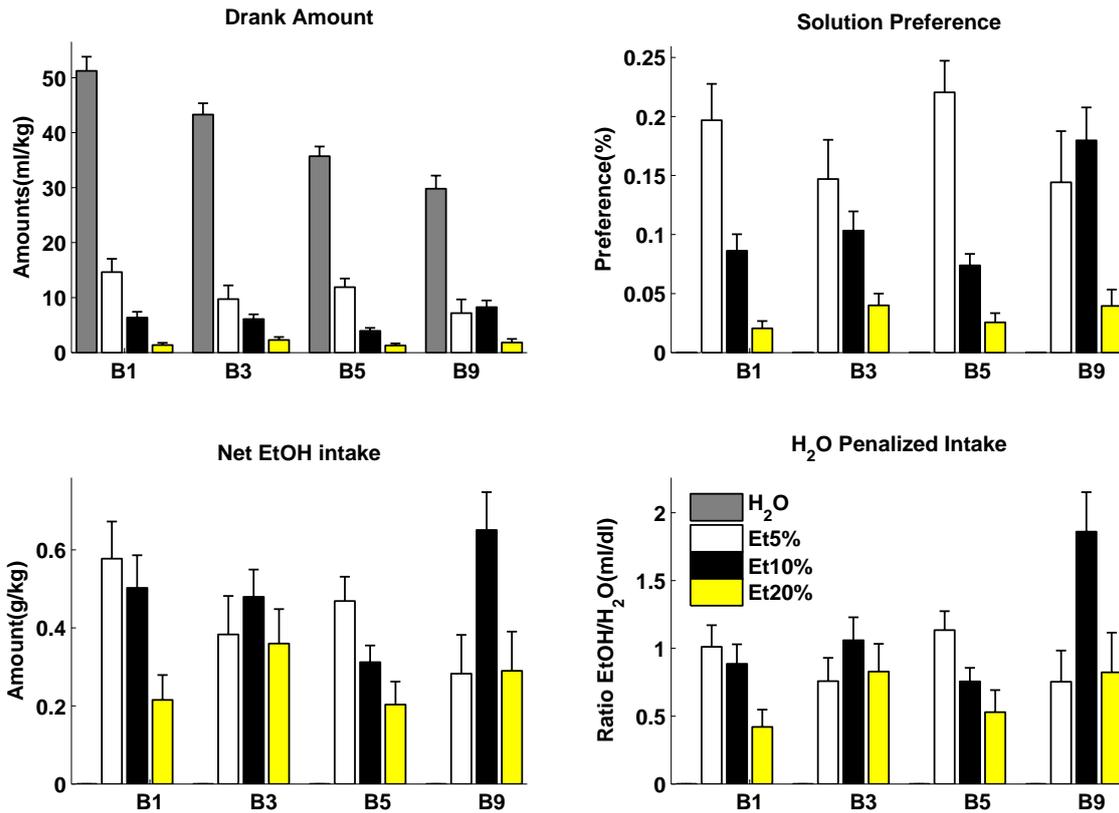
Figure 3.3: Each of the described representations gives some meaningful information while lacking some other. The best does not exist, only the most adequate given the situation.
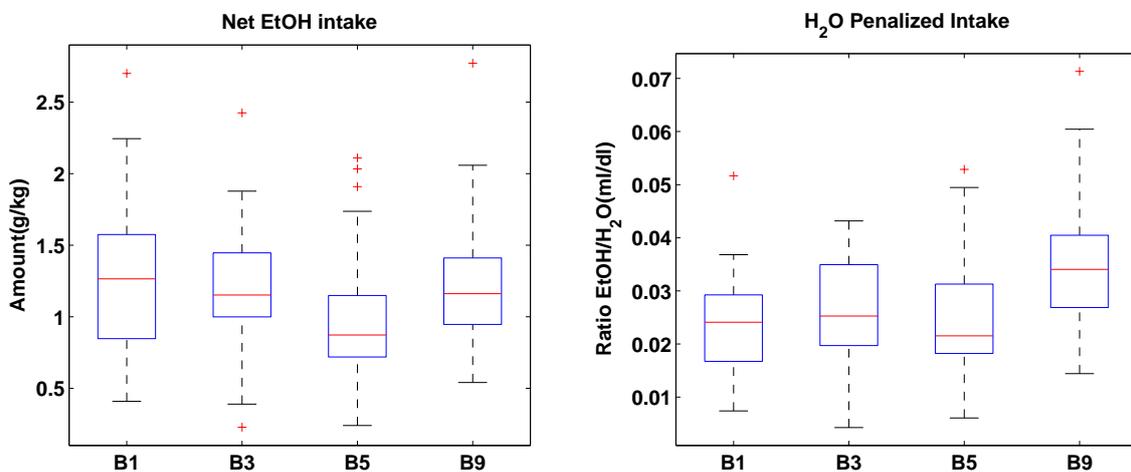
Figure 3.4: Net alcohol intake in *g* per body *kg* (left) is a standard intake measure. However, it tends to decrease for the same group of animals (B1, B3 and B5). It is important to notice that this decrease is affected by the fact that the rats gain weight throughout phases as well as they decrease drastically their water intake. H_2O penalized intake takes into account the *ml* of net ethanol consumption per *dl* of water, so that it can be observed that throughout phases, this remains constant. B9 animals come from a different experiment, and show an increased H_2O penalized intake with respect to the rest of the individuals.

3.2.2 Net alcohol preference

- Given the amount drank of each solution $EtOH_p\ p = \{.05, .1, .2\}$, the net $EtOH$ preference is calculated as

$$Pref(netEtOH_{p\%}) = \frac{p * Am(EtOH_{p\%})}{\sum_{p' \in \{.05,.1,.2\}} p' * Am(EtOH_{p'\%})} \tag{3.4}$$

- *Equation (3.4) shows the fraction of the net alcohol intake provided through each solution. It does not show how much water was simultaneously drank.*

3.2.3 Water penalized $EtOH$ intake

- Given the amount drank of each solution $EtOH_p\ p = \{.05, .1, .2\}$, the H_2O penalized $EtOH$ intake is calculated as

$$ratio(EtOH_{p\%}) = \frac{p * Am(EtOH_{p\%})}{Am(H_2O) + \sum_{p' \in \{.05,.1,.2\}} (1 - p') * Am(EtOH_{p'\%})} \tag{3.5}$$

- *Equation (3.5) shows the fraction of the net alcohol intake against H_2O provided through each solution. Penalizes hard drinkers that drink simultaneously much H_2O, e.g. first baseline.*

3.2.4 Example of different intake representations

Suppose we have 3 drinkers, each drinking only either beer, wine or spirits.

In a first example, each of them drinks the same amount of water, say 250 ml, and 15 ml of pure alcohol within the same time. Table 3.1 shows how the different drinkers would look like given different intake representations: a beer drinker has a preference for alcohol of over 50% while the spirit drinker only of 17%! The water penalized intake tells us, that the beer drinkers had 2.8 ml of pure ethanol per every liter of water while the spirit drinkers had 5,5 ml of pure ethanol per liter of water (almost twice as much!!).

Type of drinker	net $EtOH$	Am $EtOH$	H_2O	$EtOH$ Pref.	Water-pen. intake
Beer (5%)	15 ml	300 ml	250 ml	54, 55%	2,8 ml/l
Wine (12%)	15 ml	125 ml	250 ml	33, 33%	4,17ml/l
Spirit (40%)	15 ml	37,5 ml	250 ml	17, 40%	5,5ml/l

Table 3.1: Example 1 : What different $EtOH$ intake representations tell about different type of drinkers: 250 ml of pure water and 15 ml of pure ethanol drank in 300 ml, 125 ml and 37,5 ml of 5%, 12% and 40% alcoholic solutions respectively

In a second example, each of them drinks the same overall amount of liquid (water and/or alcoholic solutions), say 300 ml, and of them, 15 ml of pure alcohol, within the same time. Table 3.2 shows how the different drinkers would look like given different intake representations: a beer drinker has a preference for alcohol of 100% while the spirit drinker only of 17%. The

water penalized intake tells us, that the beer drinker consumed as much as the spirit and wine drinkers: 50 ml of pure ethanol per every liter of water.

Type of drinker	net $EtOH$	Am $EtOH$	H_2O	$EtOH$ Pref.	Water-pen. intake
Beer (5%)	15 ml	300 ml	0 ml	100%	50 ml/l
Wine (12%)	15 ml	125 ml	175 ml	41, 67%	50 ml/l
Spirit (40%)	15 ml	37,5 ml	262,5 ml	16, 7%	50 ml/l

Table 3.2: Example 2 : What different $EtOH$ intake representations tell about different type of drinkers: 300 ml of liquid and of them, 15 ml of pure ethanol consumed by means of 300 ml, 125 ml and 37,5 ml of 5%, 12% and 40% alcoholic solutions and 0 ml, 175 ml and 262,5 ml of H_2O, respectively

3.3   Individual ADE classification criterion

3.3.1   Intake increase measure for ADE criterion

For classifying an animal into presenting ADE or not, the following assumptions are made:

- Wild rats, particularly Wistar rats, present in absence of quinine ADE. This has been widely studied and results have been presented in [SS67, SSJ73, SH99, SHA$^+$96] amongst many others.

- Rats presenting after a deprivation phase an increased quinine-taste-adulterated ethanol intake, which is comparable to the increased intake of control rats, are classified as presenting ADE.

At this point, it is necessary to find a measure of the intake that allows the comparison between quinine and control groups. Figure 3.5 shows how the net $EtOH$ intake (g/kg) separates control and quinine groups, so they can not be compared at all (quinine animals present a much lower intake than controls). The H_2O penalized intake from quinine advanced drinkers (5th and 9th phases) has a "similar" distribution as in control animals. The 1st phase quinine drinkers can however be easily differentiated from controls. The latter corroborates the fact that young drinkers are susceptible to taste adulteration, thus not showing signs of addiction.

Since we are dealing with different data sets, it is also of importance that the increase from different data sets are comparable. Figure 3.6 shows (top) the net $EtOH$ increase from ($1^{st} - 5^{th}$) and 9th differ considerably; (bottom) the H_2O penalized intake behaves similarly for both data sets.

3.3.2   ADE classification criterion based on water penalized intake

Assuming that
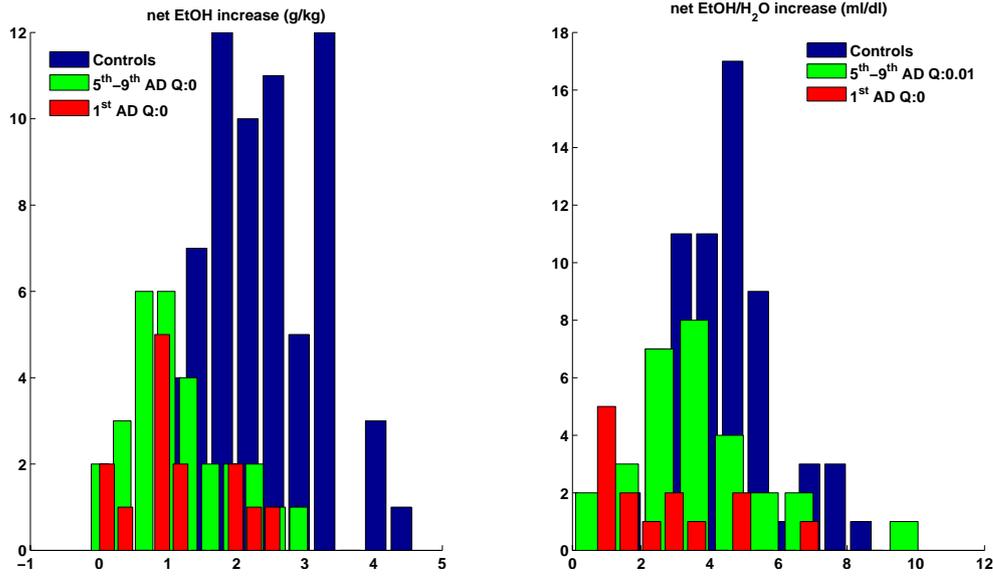
Figure 3.5: Histogram of mean increase from the first two AD days with respect to baseline. Net $EtOH$ (g/kg) increase (left): controls (blue) and quinine animals (green and red) are not comparable. H_2O penalized intake increase (right) : controls and quinine animals from 5th and 9th (advanced drinking) phases (green) are comparable (p-value = 0.01); quinine animals from the 1st phase(red) and controls cannot be compared. p-values are calculated from a two-sample KS-test.
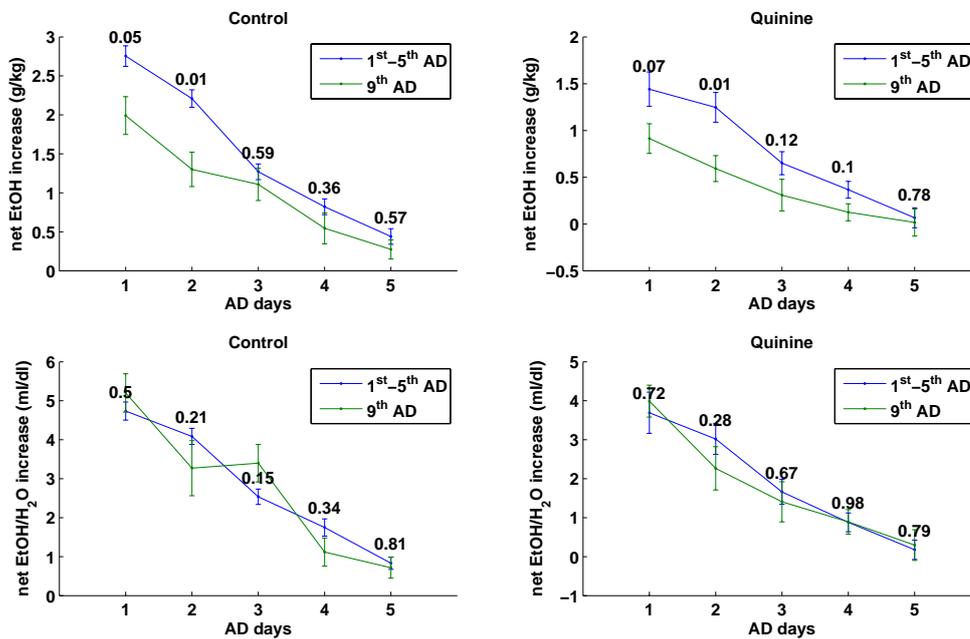
Figure 3.6: Net $EtOH$ intake (top) differentiates between the increase in the first group of animals ($1^{st} - 5^{th}$ phases) and the second group (9th phase). H_2O penalized intake, however, shows a similar intake for both controls (left) and quinine (right) animals from both data sets. P-values are computed from a two-sample t-test.

- on each day, the mean H_2O *penalized intake increase* I_d distributes $N(\mu_d, \sigma_d^2)$, $d = \{0, 1, 2, \ldots\}$, and

- during the first 2 AD days, the increase is most significant for stating ADE,

each rat r will be represented with the mean H_2O *penalized EtOH increase* on the first two AD days:

$$I^r = \frac{I_1^r + I_2^r}{2} \sim N(\frac{\mu_1 + \mu_2}{2}, \frac{\sigma_1^2 + \sigma_2^2}{2}) \tag{3.6}$$

Figure 3.3.2 plots the increase I^r as a function of the mean daily baseline H_2O penalized *EtOH* intake. No trend can be seen, which corroborates Equation (3.6).

A maximum likelihood estimator for $\hat{\mu}_C$, $\hat{\sigma}_C^2$ is obtained, fitting $\{I^{rc}\}_{rc \in \ controls}$ to the model (3.6). The likelihood of each animal (quinine and controls) given the parameters fitted for the controls $\hat{\mu}_C$, $\hat{\sigma}_C^2$ is computed as

$$L(r|\hat{\mu}_C, \hat{\sigma}_C^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}_C^2}} e^{-(x-\hat{\mu}_C)/(2\hat{\sigma}_C^2)}.$$

The animals are classified into presenting ADE (1) or not (0) by:

$$ADE(r) = \begin{cases} 1 & \text{if } I^r > \mu_C \text{ or } L(r|\hat{\mu}_C, \hat{\sigma}_C^2) > thresh \\ 0 & o.w. \end{cases}$$

3.3.3 Thresholding

Since we assume that all animals presenting ADE have $I^r \sim N(\hat{mu}_C, \hat{\sigma}_C^2)$, the threshold is selected to maximize the likelihood of the ADE classified group, given the fitted parameters for the control group.

Defining a grid $G = \{0 : 0.01 : 0.05\}$, for each threshold value $thresh \in G$ the p-value of the KS-test was computed to inspect the H_0 hypothesis: $\{I^r\}_{r:L(r|\hat{mu}_C, \hat{\sigma}_C^2) > thresh} \sim N(\hat{mu}_C, \hat{\sigma}_C^2)$. The threshold was then selected to maximize the corresponding p-value. For our data set we obtained $thresh = 0.025$. Figure 3.9 shows the threshold procedure applied to our data.

3.4 Validating the classification

A transition matrix was computed to observe differences between AD phases and the resulting classification. Each row of such matrices represents the percentage of animals from a particular phase that presented ADE or not. Figure 3.9 (left) shows no difference amongst controls throughout phases, since most of the rats presented ADE (as expected); this is also asserted through Fisher's exact test for contingency tables yielding a very high p-value (0.55). Figure 3.9 (right) shows how control animals differ in their ADE classification, according to the AD phase(p-value = 0.03).
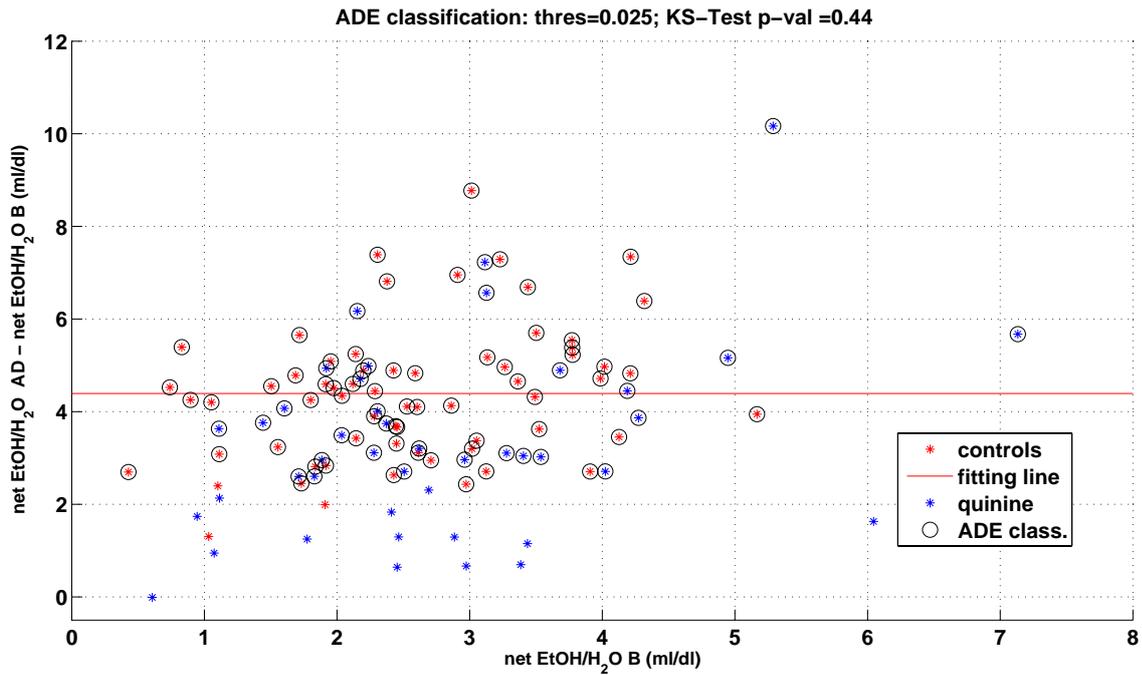
Figure 3.7: ADE classification: Mean H_2O-penalized $EtOH$ increase of the first 2 AD days (y-axis) as a function of the H_2O-penalized $EtOH$ intake (x-axis). No trend can be observed, so the increase can be modelled as a normally distributed random variable. A normal distribution (whose mean is depicted as a red line) is fitted to the controls, and the likelihood of all animals given the fitted parameters is computed. Animals presenting ADE are selected to have a likelihood greater than a threshold value (circles).
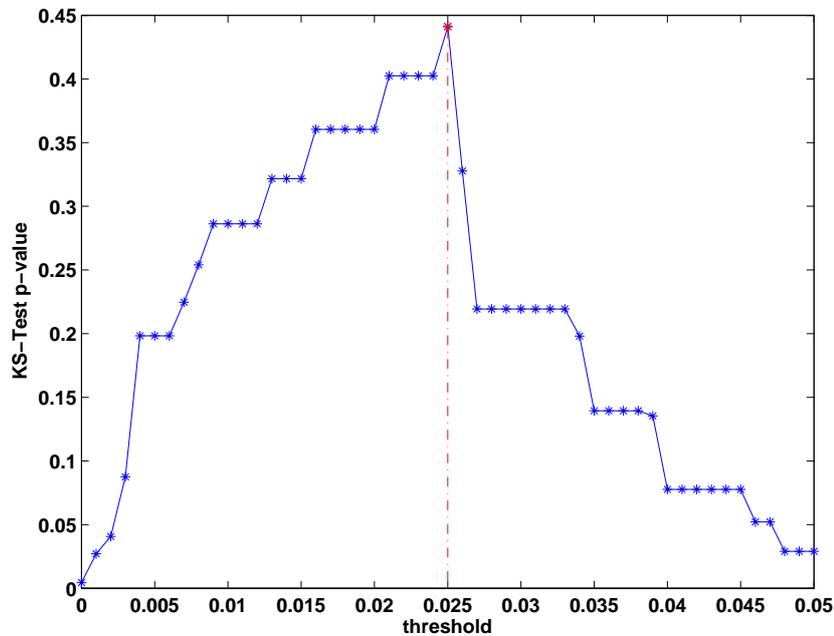
Figure 3.8: Threshold selection procedure
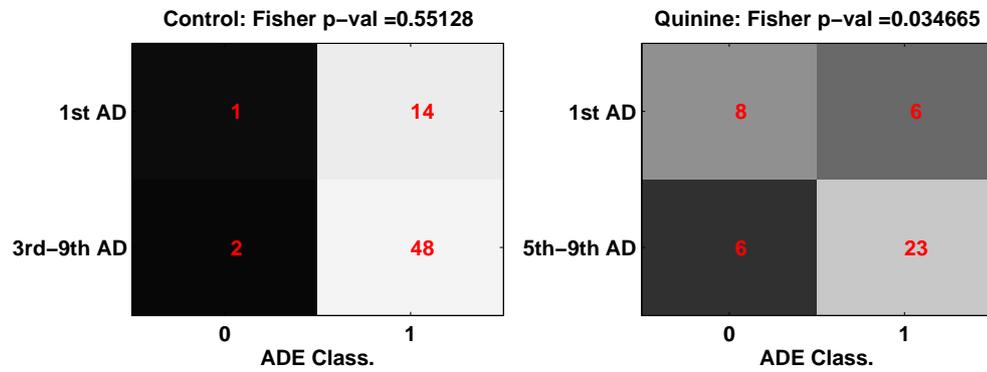
Figure 3.9: ADE classification for 1st baseline phase (top) and for the further phases (bottom) for controls (left) and quinine (right). While there is no difference for the controls (p-value=0.55), quinine animals from the first phase and those from advance AD phases are significantly different (p-value = 0.03).
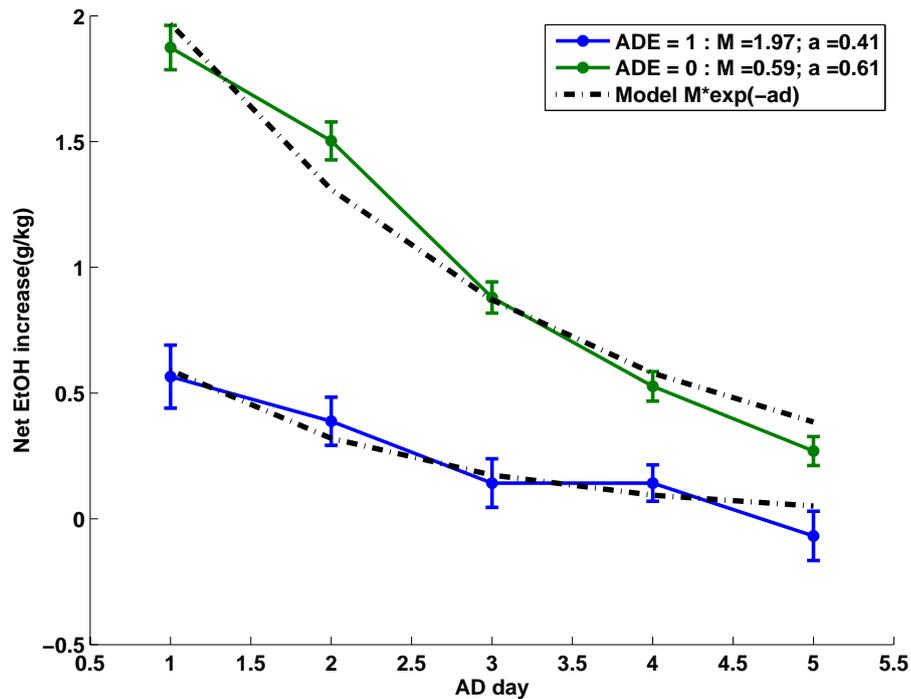
Figure 3.10: Sinclair's model fitted to ADE and non ADE animals: the model constant for ADE animals $M = 1.97$ and a $= 0.41$ are very similar to those reported by Sinclair in [SS67].

In order to compare our results with those presented in [SS67], model 3.1 was fitted to the net *EtOH* intake increase (g/kg) of both $ADE = 1$ and $ADE = 0$ groups. Results are displayed in Figure 3.10. For the $ADE = 1$ group, the fitted values $\hat{M} = 1.97$ and $\hat{a} = 0.41$ are very similar to those reported by Sinclair et al. of $M = 1.9$ and $a = 0.40$.

Chapter 4

Generalized linear modelling of time series

Generalized linear models (GLM) are a very powerful tool for modelling data assumed to be drawn from a member of the *family of exponential distributions* (FE).

The FE groups in its definition many of the most commonly used distributions. Although first defined around 1935, its convenient properties for the development of GLM has attracted much attention, so that nowadays it is still an active field of research. A very good review on its definition, moments and some example members is provided in [Cla05].

The term GLM was first defined by Nelder and Wedderburn [NW72] in 1972 as a model for random variables following a member of the FE [And70]. It provides a unifying framework for the modelling of random responses linked with systematic effects in a non-normal way. A detailed description of the model definition, iterative parameter fitting algorithm and starting approximation, several generalizations and example applications can be found in [MN89].

A generalization of GLM for expressing the existence of several latent classes is presented by Wedel and DeSarbo [WD95] with the help of the definition of mixed models and estimation maximization (EM) [DLR77]. They show furthermore how several of the *latent-class regression methods* can be seen as special cases of their proposed model. A study is developed and conclusions are drawn regarding goodness of fit and model selection as well as a wide range of example applications.

In the multivariate analysis, GLM has a key role since several of the known multivariate distributions can be also considered a member of the FE (e.g. multinomial, multivariate normal) [FT01].

4.1   Family of exponential distributions (FE)

GLM are defined for random variables Y following a member of the *natural family of exponentials* distributions. Such distributions can be parameterized in the natural form:

$$f(y|\theta, \lambda, \omega) = exp\left\{\frac{\theta'y - b(\theta)}{a(\lambda)}\omega + c(y, \lambda)\right\} \tag{4.1}$$

where a, b, c are well known functions according to the specific family, and $\theta \in \mathbb{R}^P$ is called the natural parameter. The scalar value ω is a weight used for different purposes, e.g. for grouped observations due to the same covariate values [FT01], or for clustering [WD95], as will be seen in the following chapters. The value λ is called the dispersion parameter.

By means of a simple transformation τ, (4.1) can be expressed in terms of the mean μ, where $\mu = \tau(\theta)$ is known as the *mean value natural form*. The function $\theta = \tau^{-1}(\mu)$ is called the canonical link and has important properties for the inference.

Important results on FE show that the variance structure can uniquely identify each distribution from the FE. It can also be seen that for the univariate case:

$$E[Y, \theta] = \mu = b'(\theta), V[Y, \theta] = \frac{a(\lambda)V(\mu)}{\omega} = \frac{a(\lambda)b''(\theta)}{\omega} \tag{4.2}$$

while for the multivariate:

$$E[Y, \theta] = \mu = \nabla b(\theta), Cov[Y, \theta] = \frac{a(\lambda)H_b(\theta)}{\omega} = \frac{a(\lambda)\Sigma(\mu)}{\omega} \tag{4.3}$$

with ∇b and H_b the gradient and Hessian of the multivariate function $b(\theta)$. These expressions prove to be very helpful in the statistical inference for GLM.

4.2   Generalized linear models for univariate time series

Let $y_i = \{y_{it} \in \mathbb{M}\}_{t=1}^T$, $i = 1 : N$, $\mathbb{M} \subset \mathbb{R}$ be N univariate time series consisting each of T time points. In the definition of a GLM, the three following aspects are to be taken into account:

i. The observed data y_{it} comes from a member of the family of exponential distributions, i.e. the density function has the form

$$f(y_{it}|\theta_{it}, \lambda_{it}, \omega_{it}) = \exp\left\{\frac{y_{it}\theta_{it} - b(\theta_{it})}{a(\lambda)}w_{it} + c(y_{it}, \lambda)\right\}. \tag{4.4}$$

ii. The covariate vector $x_{it} \in \mathbb{R}^P$ influences y_{it} in the form of a linear predictor

$$\eta_{it} = Z_{it}\beta, \ \beta \in \mathbb{R}^P, \ Z_{it} = Z(x_{it}'), \tag{4.5}$$

where Z_{it} is a work matrix given for flexibility. In the univariate case, Z_{it} can be directly defined as $Z_{it} = x_{it}'$.

iii. The linear predictor is related to the mean μ_{it} through the *response function* $h : \mathbb{R} \to \mathbb{M}$, i.e.

$$\mu_{it} = h(\eta_{it}). \tag{4.6}$$

If the inverse function $g = h^{-1} : \mathbb{M} \to \mathbb{R}$ exists then

$$g(\mu_{it}) = \eta_{it} = Z_{it}\beta, \tag{4.7}$$

where g is then called the *link function*.

A *canonical link* occurs when $g = \tau^{-1}$, such that $\theta_{it} = \eta_{it}$, which yields a simplification of the fitting algorithms and provides a set of sufficient statistics [NW72]. Given a sufficient statistic $T(X)$ for a parameter θ of a random variable X, the inference of θ can be made based on the statistic, since the sample does not provide further information as the sample itself. A characterization of $T(X)$ is given by the Fisher's factorization theorem stating that $T(X)$ can be decomposed into

$$T(X) = h(X)g_\theta(T(X)),$$

so that in the estimation of θ, only $T(X)$ can be used.[1]

4.2.1 Maximum likelihood estimation

The model parameter vector $\beta = \{\beta_p\}_{p=1}^P$ is estimated as the maximizer of the log-likelihood function:

$$L = log f(y_1, \ldots, y_N | \theta_1, \ldots, \theta_N, \lambda, \omega_1, \ldots, \omega_N). \tag{4.8}$$

Conditional on the covariates, the observations are assumed to be independent within individuals and time points. To consider the important information given by the time structure (otherwise ignored by the independence assumption) the covariates are properly selected to express the temporal structure, e.g. trigonometric polynomials representing seasonal components or monotonic functions of it to include trends, etc. Equation (4.8) reduces under this assumption to:

$$
\begin{aligned}
L(y|\beta, \omega) &= log(\prod_{i=1}^N \prod_{t=1}^T f(y_{it}|\theta_{it}, \lambda, \omega_{it}) \\
&= \sum_{i=1}^N \sum_{t=1}^T log f(y_{it}|\theta_{it}, \lambda, \omega_{it})
\end{aligned}
$$

$$L(y|\beta, \omega) = \sum_{i=1}^N \sum_{t=1}^T \left\{ \frac{y_{it}\theta_{it} - b(\theta_{it})}{a(\lambda)}\omega_{it} + c(y_{it}, \lambda) \right\}. \tag{4.9}$$

Applying the necessary condition of extrema:

$$\beta^* = \arg\max_{\beta \in \mathbb{R}^P} L(y) \Rightarrow \frac{\partial L}{\partial \beta_p}(y, \beta^*) = 0$$

---

[1]The short description of the sufficient statistic was adapted from http://en.wikipedia.org/wiki/Sufficient_statistic

the homogenous system of nonlinear equations

$$\frac{\partial L}{\partial \beta_p}(y, \beta^*) = \sum_{i,t=1}^{N,T} \frac{\partial L}{\partial \theta_{it}} \frac{\partial \theta_{it}}{\partial \mu_{it}} \frac{\partial \mu_{it}}{\partial \eta_{it}} \frac{\partial \eta_{it}}{\partial \beta_p} = 0$$

is obtained, whose solution is a local maximum of the likelihood function. As shown in [NW72]:

$$b'(\theta_{it}) = \mu_{it}, \quad b''(\theta_{it}) = \frac{\partial \mu_{it}}{\partial \theta_{it}} = \frac{var(y_{it})}{a(\lambda)} \omega_{it} = V_{it}(\mu) \quad \text{and } \eta_{it} = \sum_{p=1}^{P} x_{itp} * \beta_p,$$

so that

$$\frac{\partial L}{\partial \beta_p}(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} \frac{y_{it} - \mu_{it}}{V_{it}} \frac{\partial \mu_{it}}{\partial \eta_{it}} x_{itp} = 0. \tag{4.10}$$

To solve the homogeneous nonlinear system of equations $(\nabla_L(\beta) = 0)$ given by (4.10), a Newton-Raphson method is developed so that:

$$-H_L(\beta)(x_{k+1} - x_k) = \nabla_L(\beta), \tag{4.11}$$

where $H_L(\beta) = (H_{p,q})_{p,q=1}^{P}$ is the matrix of the second derivatives of L with respect to β:

$$
\begin{aligned}
H_L(p,q) &= \frac{\partial^2 L}{\partial \beta_p \beta_q} \\
&= \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} \frac{\partial}{\partial \eta_{it}} \left(\frac{y_{it} - \mu_{it}}{V_{it}} \frac{\partial \mu_{it}}{\partial \eta_{it}} x_{itp} \right) \frac{\partial \eta_{it}}{\partial \beta_q} \\
&= \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} \left[\frac{-1}{V_{it}} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}} \right)^2 + \frac{y_{it} - \mu_{it}}{V_{it}} \frac{\partial^2 \mu_{it}}{\partial \eta_{it}^2} \right] x_{itp} x_{itq}.
\end{aligned}
$$

Instead of H_L, its expected value $\bar{H}_L$ is used to solve the system. This procedure is known as the *Fisher's method of scoring* [Osb92]. Then:

$$\bar{H}_L(p,q) = -\frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} w_{it} x_{itp} x_{itq},$$

where $w_{it} = \omega_{it} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}} \right)^2 / V_{it}$. In a matrix form this can be seen as

$$\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1}^{N} x_i' diag(w_i) x_i, \tag{4.12}$$

with x_i the $T * P$ matrix of covariates associated with the ith time series.

Following the same notation, (4.10) can be represented as:

$$\frac{\partial L}{\partial \beta_p}(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i=1}^{N} x_i' * diag(w_i) diag(\frac{\partial \eta_i}{\partial \mu_i})(y_i - \mu_i) = 0. \tag{4.13}$$

Equations (4.13), (4.12) together with (4.11) can be used to iteratively obtain a good approximation of the maximum likelihood estimator β^*.

4.2.2 Canonical link

When $\theta = \eta$ equations (4.13), (4.12) reduce to:

$$\frac{\partial L}{\partial \beta_p}(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i=1}^{N} x_i' diag(\omega_i)(y_i - \mu_i) = 0, \tag{4.14}$$

$$\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1}^{N} x_i' diag(\omega_i) diag(V_i) x_i. \tag{4.15}$$

Table 4.1 shows the expressions of L, V, $\mu(\eta)$, $V(\mu)$ and $a(\lambda)$ of several known members of the FE, which can be used directly in the programming of the described algorithms given a canonical link.

Distribution	Notation	$L(y, \mu)$	$\mu(\eta)$	$V(\mu)$	$a(\lambda)$
Normal	$N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\frac{y-\mu)^2}{2\sigma^2})}$	η	1	σ^2
Binomial	$B(p = \mu/n, n)$	$\binom{n}{y} p^y (1-p)^{(n-y)}$	$n\frac{e^\eta}{1+e^\eta}$	$\mu(n-\mu)$	1
Poisson	$P(\mu)$	$\frac{e^{-\mu}\mu^y}{y!}$	e^η	μ	1
Gamma	$\Gamma(k, \mu)$	$\left(\frac{k}{\mu}\right)^k y^{k-1}\frac{e^{-ky/\mu}}{\Gamma(k)}$	$-1/\eta$	μ^2	k^{-1}

Table 4.1: GLM elements for some known distributions.

4.3 Generalized linear models for multivariate time series

Let now $y_i = \{y_{it}\}_{t=1}^T$, with $y_{it} = \{y_{it1}, \ldots, y_{itR}\}$, so that at each time point t, an individual i has a multivariate response of length R. In this case, y_{it} is assumed to be the realization of a multivariate variable from the multivariate family of exponentials, amongst which the multinomial and the multivariate normal can be considered.

As in the univariate case, the three elements of a GLM have to be taken into account, i.e.

i. The observed data $y_{it} \in \mathbb{M}^R$ comes from a multivariate exponential family of distributions with R response channels, i.e. the density function has the form

$$f(y_{it}|\theta_{it}, \lambda_{it}, \omega_{it}) = \exp\left\{\frac{y_{it}'\theta_{it} - b(\theta_{it})}{a(\lambda)}w_{it} + c(y_{it}, \lambda)\right\}. \tag{4.16}$$

In this case $b : \mathbb{R}^R \to \mathbb{R}$.

ii. The covariate vector $x_{it} \in \mathbb{R}^P$ influences y_{it} in the form of a linear predictor

$$\eta_{it} = Z_{it}\beta, \ \beta \in \mathbb{R}^{PR}, \ Z_{it} = Z(x'_{it}) = blockdiag\left([x'_{it}]_{r=1}^R\right). \tag{4.17}$$

The expression of the work matrix Z_{it} will be better described later in this chapter.

iii. The linear predictor is related to the mean μ_{it} through the *response function* $h : \mathbb{R}^R \to \mathbb{M}^R$, i.e.

$$\mu_{it} = h(\eta_{it}). \tag{4.18}$$

If the inverse function $g = h^{-1} : \mathbb{M}^R \to \mathbb{R}^R$ exists then

$$g(\mu_{it}) = \eta_{it} = Z_{it}\beta \tag{4.19}$$

is the *link function*.

4.3.1   Maximum likelihood estimation

The model parameter vector $\beta = \{\beta_p\}_{p=1}^{PR}$ is estimated as the maximizer of the log-likelihood function, as sketched in Section 4.2.1, making the corresponding changes for the multivariate analysis. Applying the necessary condition of maxima

$$\beta^* = \arg\max_{\beta \in \mathbb{R}^{PR}} L(y) \Rightarrow \frac{\partial L}{\partial \beta}(y, \beta^*) = 0,$$

we obtain

$$
\begin{aligned}
\frac{\partial L}{\partial \beta}(y, \beta^*) &= \sum_{i,t=1}^{N,T} \frac{\partial L}{\partial \theta_{it}} \frac{\partial \theta_{it}}{\partial \mu_{it}} \frac{\partial \mu_{it}}{\partial \eta_{it}} \frac{\partial \eta_{it}}{\partial \beta_{pr}} \\
&= \sum_{i,t=1}^{N,T} \frac{\omega_{it}}{a(\lambda)} \left(y'_{it} - \frac{\partial b(\theta_{it})}{\partial \theta_{it}}\right) \frac{\partial \mu_{it}}{\partial \theta_{it}}^{-1} \frac{\partial \mu_{it}}{\partial \eta_{it}} \frac{\partial \eta_{it}}{\partial \beta} \\
&= 0.
\end{aligned}
$$

As shown in Section 4.1

$$
\begin{aligned}
\frac{\partial b(\theta_{it})}{\partial \theta_{it}} &= \mu'_{it}, \\
\frac{\partial \mu_{it}}{\partial \theta_{it}} &= \frac{\partial^2 b(\theta_{it})}{\partial \theta_{it} \partial \theta_{it}} \\
&= \frac{Cov[Y, \theta]}{a(\lambda)} \omega_{it} = \Sigma_{it}(\mu) \\
\eta_{it} &= Z_{it}\beta.
\end{aligned}
$$

The work matrix $Z_{it}(x'_{it}) \in \mathbb{R}^{R \times RP}$ is a block diagonal matrix containing as many rows as response channels of the modelled multivariate variable, and in each row the covariate vector x'_{it} in the columns from $(r-1) * P + 1$ to rP, as expressed in the equation

$$Z_{it}(r, q) = \begin{cases} x_{itp}, & \text{if } q \in \mathbb{Z} \bigcap [(r-1)P + 1, rP] \text{ and } p = mod(q, P) + 1 \\ 0, & \text{o.w.} \end{cases}$$

Z_{it} has the following portrait:

$$\begin{bmatrix} x'_{it} & 0 & \dots & 0 \\ 0 & x'_{it} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & x'_{it} \end{bmatrix}.$$

The necessary condition for extrema takes the form:

$$\frac{\partial L}{\partial \beta}(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it}(y'_{it} - \mu'_{it})\Sigma_{it}(\mu_{it})^{-1}\frac{\partial \mu_{it}}{\partial \eta_{it}} Z_{it} = 0. \tag{4.20}$$

A Newton-Raphson method is again employed to solve the system of nonlinear equations $\nabla_L(\beta) = \left(\frac{\partial L}{\partial \beta}\right)' = 0$ given by (4.20) so that

$$-H_L(\beta)(x_{k+1} - x_k) = \nabla_L(\beta), \tag{4.21}$$

where $H_L(\beta) = (H_{p,q})_{p,q=1}^P$ is the matrix of the second derivatives of L with respect to β, i.e.

$$\begin{aligned} H_L &= \frac{\partial^2 L}{\partial \beta \partial \beta} \\ &= \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} Z'_{it} \frac{\partial}{\partial \beta} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}'} \Sigma_{it}(\mu_{it})^{-1}(y_{it} - \mu_{it}) \right) \\ &= \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} Z'_{it} \left[\frac{\partial}{\partial \beta} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}'} \Sigma_{it}(\mu_{it})^{-1} \right) (y_{it} - \mu_{it}) + \left(\frac{\partial \mu_{it}}{\partial \eta_{it}'} \Sigma_{it}(\mu_{it})^{-1} \right) \frac{\partial}{\partial \beta} (y_{it} - \mu_{it}) \right]. \end{aligned}$$

Using H_L instead of its expected value $\bar{H}_L$, defined by

$$\begin{aligned} \bar{H}_L(p,q) &= \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} Z'_{it} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}'} \Sigma_{it}(\mu_{it})^{-1} \right) \frac{\partial}{\partial \mu}(y_{it} - \mu_{it}) \frac{\partial \mu_{it}}{\partial \eta_{it}} \frac{\partial \eta_{it}}{\partial \beta} \\ &= -\frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} \omega_{it} Z'_{it} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}'} \Sigma_{it}(\mu_{it})^{-1} \frac{\partial \mu_{it}}{\partial \eta_{it}} Z_{it} \right), \end{aligned}$$

and denoting

$$W_{it} = \left(\frac{\partial \mu_{it}}{\partial \eta_{it}'} \Sigma_{it}(\mu_{it})^{-1} \frac{\partial \mu_{it}}{\partial \eta_{it}} \right) \text{ and } \Omega_{it} = diag(\omega_{it}e_R), \text{ where } e_R \in \mathbb{R}^R, \; e_R(r) = 1, \; r = 1, \dots, R,$$

we have

$$\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1,t=1}^{N,T} Z'_{it} W_{it} \Omega_{it} Z_{it}. \tag{4.22}$$

Following the same notation, (4.20) can be represented as:

$$\frac{\partial L}{\partial \beta}(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i,t=1}^{N,T} Z'_{it} W_{it} \frac{\partial \eta_{it}}{\partial \mu_{it}} \Omega_{it}(y_{it} - \mu_{it}) = 0. \tag{4.23}$$

Equations (4.23), (4.22) together with (4.21) can be used to iteratively obtain a good approximation of the maximum likelihood estimator β^*.

Constructing the N matrices $Z_i \in \mathbb{R}^{RP \times RT}$:

$$Z'_i = [Z'_{i1}, Z'_{i2}, \ldots, Z'_{iT}],$$

the block-wise diagonal matrices

$$W_i = blockdiag([W_{it}]_{t=1}^T),$$

$$\Omega_i = diag([\Omega_{it}]_{t=1}^T),$$

$$\frac{\partial \eta_i}{\partial \mu_i} = diag\left(\left[\frac{\partial \eta_{i1}}{\partial \mu_{i1}}', \frac{\partial \eta_{i2}}{\partial \mu_{i2}}', \ldots, \frac{\partial \eta_{iT}}{\partial \mu_{iT}}'\right]\right),$$

and the vectors

$$y'_i = [y'_{i1}, y'_{i2}, \ldots, y'_{iT}],$$

$$\mu'_i = [\mu'_{i1}, \mu'_{i2}, \ldots, \mu'_{iT}],$$

Equations (4.22) and (4.23) can be seen as:

$$\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1}^N Z'_i W_i \Omega_i Z_i \tag{4.24}$$

$$\nabla L(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i=1}^N Z'_i W_i \frac{\partial \eta_i}{\partial \mu_i} \Omega_i(y_i - \mu_i) = 0. \tag{4.25}$$

In order to obtain a flexible representation, allowing different input and storage formats, the permutation matrix P is defined so that $Py = y_1$, where the rows of y_1 have a different ordering as the ones of the input.

Some of the properties of the permutation matrices of interest are:

- Given a permutation vector π of the numbers $1 \ldots N$, the associated permutation matrix is defined as:
$$P(i,j) = \begin{cases} 1, & \text{if j} = \pi_i \\ 0, & \text{o.w.} \end{cases}$$

- PA permutes the rows of A so that the $i - th$ row of PA is the $\pi(i) - th$ row of A.

- AP' permutes the columns of A so that the $i-th$ column of AP' is the $\pi(i)-th$ row of A.

- $P^{-1} = P^T$

Applying the same permutation P to Equations 4.24 and 4.25, an equivalent system is obtained with matrix

$$\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1}^{N} Z_i'P'PW_iP'P\Omega_iP'PZ_i = -\frac{1}{a(\lambda)} \sum_{i=1}^{N} \tilde{Z}_i'\tilde{W}_i\tilde{\Omega}_i\tilde{Z}_i, \tag{4.26}$$

where $\tilde{Z}_i = PZ_i$, $\tilde{W}_i = PW_iP'$, $\tilde{\Omega}_i = P\Omega_iP'$, and the right hand side

$$\nabla L(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i=1}^{N} \tilde{Z}_i'\tilde{W}_i\frac{\partial\tilde{\eta}_i}{\partial\tilde{\mu}_i}\tilde{\Omega}_i(\tilde{y}_i - \tilde{\mu}_i) = 0 \tag{4.27}$$

with $\frac{\partial\tilde{\eta}_i}{\partial\tilde{\mu}_i} = P\frac{\partial\eta_i}{\partial\mu_i}P'$, which can be used to modify the fitting algorithms for different data input formats and for optimized storage.

As an example of the described modification, a permutation is proposed, such that,

$$\tilde{y}_i = Py_i = [y1_i', \ldots, yR_i']'$$

where $yr_i \in \mathbb{M}^T$ is a time series representing the $r-th$ output channel of the $i-th$ individual. The corresponding permutation vector is

$$\pi_1[(r-1)T+j] = r + (j-1)*R, \text{ For } j = 1\ldots T \text{ and } r = 1\ldots R.$$

The matrix of covariates and its transformed form for the permutation vector π are portrayed in Figure 4.1. The matrix W_i and its row and column transformed version $\tilde{W} = PW_iP'$ are also portrayed in Figure 4.2.

4.3.2 Canonical link

If a canonical link is used so that $h = \tau$, Equations (4.27) and (4.26) reduce to:

$$\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1}^{N} \tilde{Z}_i'\tilde{\Sigma}_i\tilde{\Omega}_i\tilde{Z}_i \tag{4.28}$$

and

$$\nabla L(y, \beta^*) = \frac{1}{a(\lambda)} \sum_{i=1}^{N} \tilde{Z}_i'\tilde{\Omega}_i(\tilde{y}_i - \tilde{\mu}_i) = 0, \tag{4.29}$$

where $\tilde{\Sigma}_i = P\Sigma_iP'$ and Σ_i is the block diagonal matrix $diag([\Sigma_{it}(\mu_{it})]_{t=1}^T)$ whose portraits are the same as those in Figure 4.2.
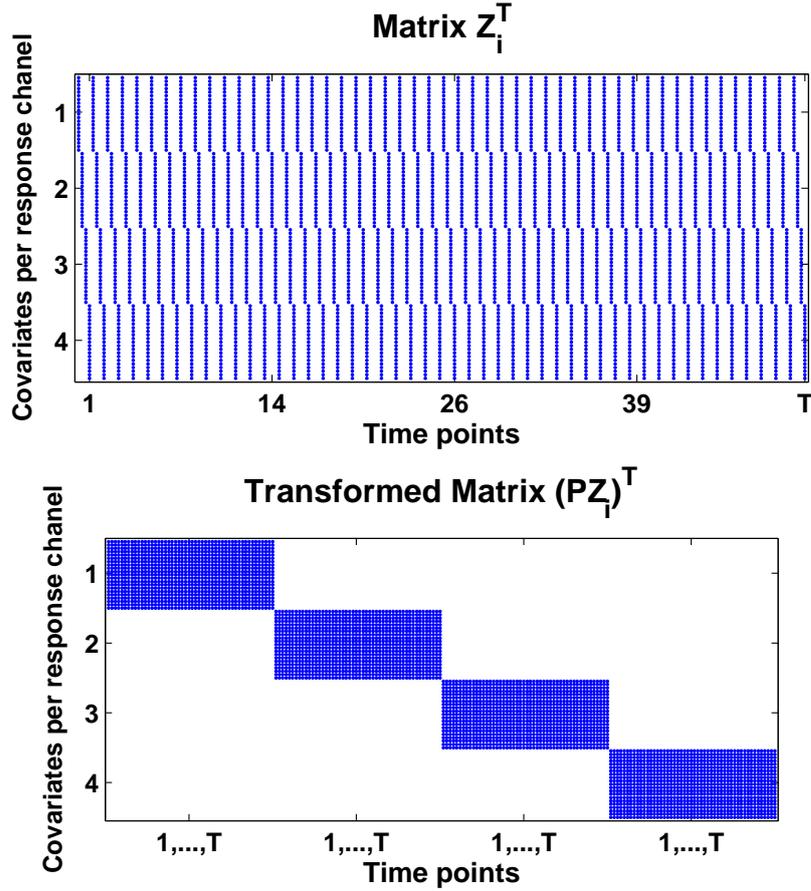
Figure 4.1: Matrix Z_i' (top) and the new matrix $\tilde{Z}_i = Z'P_i'$ after applying the permutation columns (bottom).

The estimator β for the canonical link is proved to be consistent and asymptotically normal under appropriate assumptions [FT01].

Multinomial log-it

An example of multivariate time series GLM is the multinomial distribution with its canonical link, the generalized logistic function.

The multinomial distribution is one of the multivariate family of distributions. Given

$$Y \sim multinomial(n, \pi), \text{i.e. } Y \in \{\mathbb{Z} \bigcap [0; n]\}^R,$$

where $Y(r)$ is the amount of times out of n the event r occurred, so that $\sum_{r=1}^{R+1} Y(r) = n$. The event r has a probability of occurrence π_r, so that $\sum_{r=1}^{R+1} \pi_r = 1$. The value n is known and the parameter vector $\pi = \{\pi_1; \ldots; \pi_R\}$ is to be estimated.

The multinomial density function is given by:
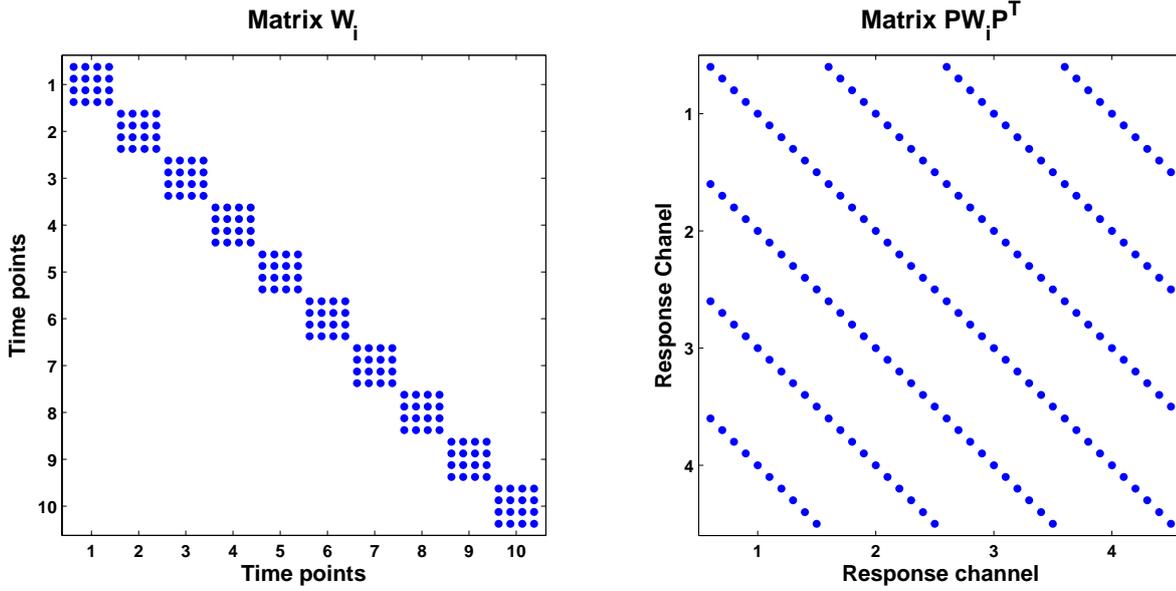
Figure 4.2: Matrix W_i (top) and the new matrix $\tilde{W}_i = PW_iP'$ after applying the permutation P to its rows and columns (bottom).

$$f(Y = y|n, \pi) = \frac{n!}{\prod_{r=1}^{R+1} y_r!} \prod_{r=1}^{R+1} \pi_r^{y_r},$$

having $\pi_{(R+1)} = 1 - \sum_{r=1}^{R} \pi_r$ and $y_{(R+1)} = n - \sum_{r=1}^{R} y_r$.

The canonical link for this distribution is the generalized logistic function:

$$\pi_r = \tau(\theta_r) = h(\eta_r) = \frac{exp(\eta_r)}{1 + \sum_{r'=1}^{R} exp(\eta_{r'})}, \ r = 1 \ldots R \text{ and } \pi_{R+1} = \frac{1}{1 + \sum_{r'=1}^{R} exp(\eta_{r'})}.$$

The covariance matrix under this distribution is:

$$\Sigma(\pi) = \begin{pmatrix} \pi_1(1 - \pi_1) & -\pi_1\pi_2 & \ldots & -\pi_1\pi_R \\ \vdots & \ddots & \ddots & \vdots \\ -\pi_R(\pi_1) & -\pi_R\pi_2 & \ldots & \pi_R(1 - \pi_R) \end{pmatrix}.$$

The matrix H_L and right hand side vector ∇_L are:

$$H_L = \sum_{i=1}^{N} \tilde{Z}_i' \tilde{\Sigma}_i \tilde{\Omega}_i \tilde{Z}_i,$$

where $\tilde{\Sigma}_i = P\Sigma_iP'$ and Σ_i is the block diagonal matrix $\Sigma_i = diag([\Sigma_{it}(\pi_{it})]_{t=1}^{T})$ and

$$\nabla_L = \sum_{i=1}^{N} \tilde{Z}_i' \tilde{\Omega}_i (\tilde{y}_i - \tilde{\mu}_i).$$

Figure 4.3 shows simulated data following a multinomial distribution with parameter $\pi_t = h(Z_t\beta)$. In the second row, the estimated model is compared to the original data set by means of the mean value of Y at time t and the expected mean value μ given the set of parameters from the fitted model. This data has a similar behavior as the real data analyzed as part of this thesis in later chapters.
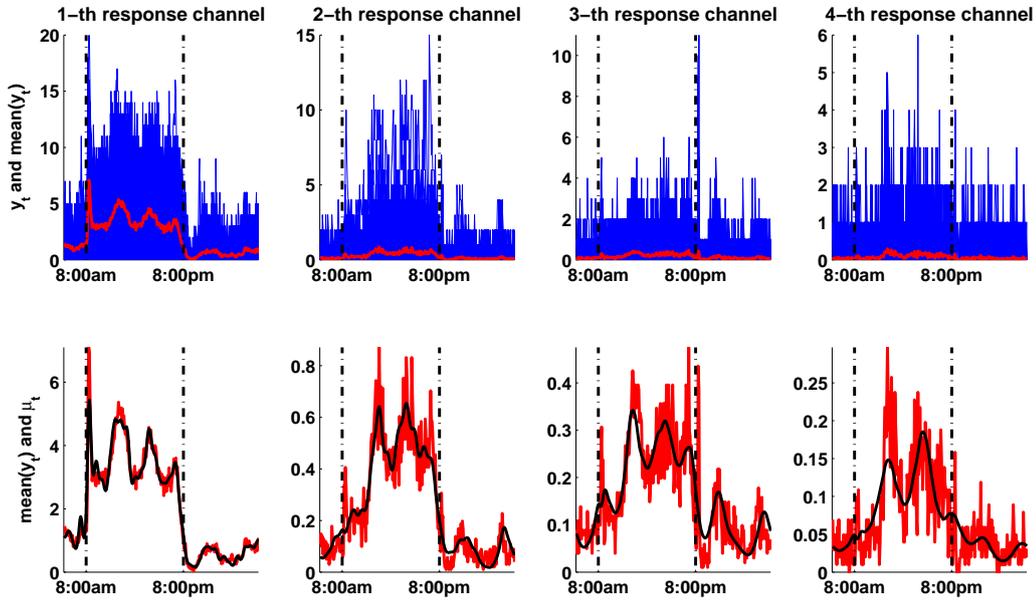
Figure 4.3: Example of multivariate data (blue) y, its mean value $mean(y)$ (red lines), and the expected value given the described model $\mu = E(Y|\beta, X) = n\pi$ (black).

4.4 Modifications for computational efficiency

Theoretically, the explained methods can be applied without problems to any data set under the mentioned distributional assumptions. However, computational resources often requiere a simplification of the conditions, in order to improve computational costs, reduce storage resources, amongst others.

The direct application of the described methods is extremely costly for bigger data sets. Only building the N matrices Z_i would requiere the storage of $PTRN$ non-zeros in the form of N sparse matrices, each of size $PR \times TR$. W_i requires NTR^2 non-zeros in a sparse $NTR \times NTR$ matrix.

In order to obtain less costly algorithms, we suggest the following modifications.

4.4.1 Constant covariates along individuals

Instead of considering for every individual a covariates matrix $x_i = (x_{itp})_{t=1,p=1}^{T,P}$, a constant set of covariates along individuals can be selected, such that $x_{it} = x_t$, $i = 1 \ldots N$. It is based on

the fact that a model of a group as a whole is desired, so that an understanding of the unifying process can be drawn. This assumption implies:

$$
\begin{aligned}
\tilde{Z}_i &= \tilde{Z} \\
\tilde{\eta}_i &= \tilde{Z}_i \beta = \tilde{Z} \beta = \tilde{\eta} \\
\tilde{\mu}_i &= h(\tilde{\eta}_i) = h(\tilde{\eta}) = \tilde{\mu} \\
\tilde{W}_i &= \tilde{W} \\
\tilde{\Sigma}_i &= \tilde{\Sigma}.
\end{aligned}
$$

Equations (4.26) and (4.27) now are:

$$
\bar{H}_L(\beta) = -\frac{1}{a(\lambda)} \sum_{i=1}^{N} \tilde{Z}' \tilde{W} \Omega_i \tilde{Z} = -\frac{1}{a(\lambda)} \tilde{Z}' \tilde{W} \sum_{i=1}^{N} \tilde{\Omega}_i \tilde{Z} \tag{4.30}
$$

and

$$
\begin{aligned}
\nabla L(y, \beta^*) &= \frac{1}{a(\lambda)} \sum_{i=1}^{N} \tilde{Z}' \tilde{W} \frac{\partial \tilde{\eta}}{\partial \tilde{\mu}} \tilde{\Omega}_i (\tilde{y}_i - \tilde{\mu}) \\
&= \frac{1}{a(\lambda)} \tilde{Z}' \tilde{W} \frac{\partial \tilde{\eta}}{\partial \tilde{\mu}} \left\{ \left[\sum_{i=1}^{N} \tilde{\Omega}_i \tilde{y}_i \right] - \left[\sum_{i=1}^{N} \tilde{\Omega}_i \right] \tilde{\mu} \right\}
\end{aligned} \tag{4.31}
$$

respectively.

The quantities $\left[\sum_{i=1}^{N} \tilde{\Omega}_i \tilde{y}_i \right]$ and $\left[\sum_{i=1}^{N} \tilde{\Omega}_i \right]$ are computed at the beginning of the algorithm and only once. Each iteration of the Newton-Raphson method requieres a single update of the matrices $\tilde{W}$, $\frac{\partial \tilde{\eta}}{\partial \tilde{\mu}}$ and $\tilde{\mu}$ for the new estimation of β, and would need to perform the expressed matrix products only once, and not N times per iteration.

This simplification achieves a tremendous storage and computing time efficiency, since:

- only a single sparse $KP \times TK$ matrix is stored (Z) with RTP non zero elements (versus N such matrices in the general case),

- at every iteration only the vector $\mu \in \mathbb{M}^T$, and the matrices $\tilde{W}$ and $\frac{\partial \tilde{\eta}}{\partial \tilde{\mu}}$ have to be computed, and

- the amount of matrices products is considerably reduced.

Of course, taking this modification into account, many of the general modelling features of GLM are lost, since those models including individual measurements, like Markov chains, can not be modelled with this approach.

4.4.2 Variable covariates amongst response channels (for the multivariate case)

In the data sets analyzed in this thesis, we observed that some of the response channels were richly featured, while other channels contained relatively few interesting components. I.e. some channels require many covariates to achieve a good mean approximation and others need very
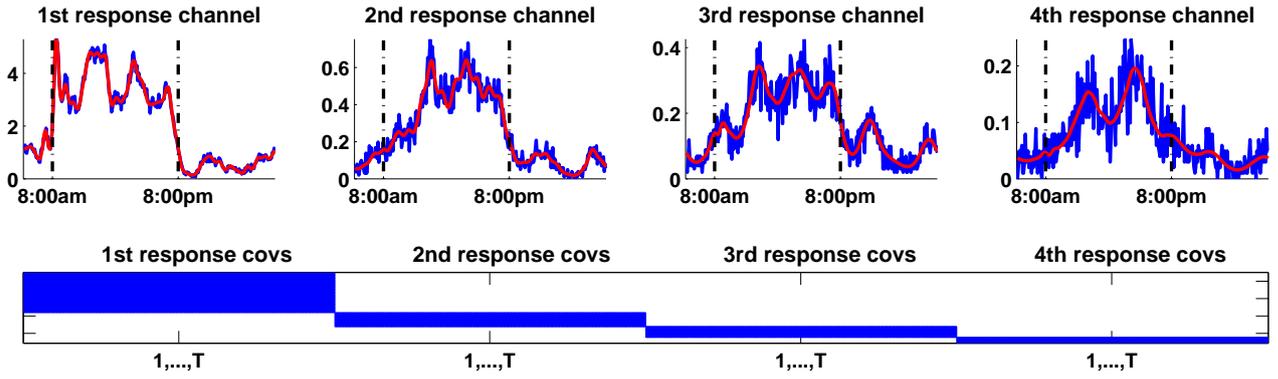
Figure 4.4: Mean value of y (blue) and estimated mean value (red) through the logit-multinomial model (top). The bottom figure shows the portrait of the matrix Z, having 37, 13, 10 and 5 covariates per response channel respectively.

few components.

The comfortable representation given by $\tilde{Z}$ allows to select for each response dimension how many, and furthermore which, covariates are of interest. For this we employed a model selection technique based on a modified Akaike information criterion (AIC) suitable for model selection of GLM [WD95]. The CAIC of a model is defined as

$$CAIC(y, \beta) = -2L + Npar * (log(Nobs) + 1),$$

with L being the log likelihood of the model, given the data y and the parameters β, $Npar$ the number of parameters (i.e. $|\beta|$), and $Nobs$ the number of observations (i.e. $|y|$).

To select the number of covariates, the general model β_0 is fitted to the data, and the $CAIC_0$ is computed. Then a new model is fitted eliminating one of the covariates, and the $CAIC_1$ is computed. If $CAIC_0 > CAIC_1$, then the mentioned covariate is selected as meaningful to the data. If, on the other hand, $CAIC_0 \leq CAIC_1$, this covariate has no mayor significance for the model and is eliminated. The same procedure is repeated with all the covariates until a subset of the original set is selected. For the remaining subset, the described procedure is repeated until the initial and final sets are equal.

Figure 4.4 shows the mean value of a data set y and its estimated mean value through the logit-multinomial model (top). As can be seen, the features of the rightmost channel can be estimated with fewer covariates than the leftmost one. The bottom figure shows the portrait of the matrix (Z), having 37, 13, 10 and 5 covariates per response channel, respectively. The covariates were selected through the described procedure.

Chapter 5

Time series GLM-based EM clustering

IN the practical applications of statistics such problems as the following often present themselves: Of n households exposed to risk, m_0 returned 0 cases of disease, m_1 returned each a single case, m_2 each two cases, ..., m_n each n cases. Might such a distribution have arisen from sampling a " population" each member of which was subject to a constant chance of infection throughout the period of exposure, or is the form of the distribution valid evidence that particular households were especially prone to take the disease in question ?

<div align="right">

Major Greenwood and G. Udny Yule

March, 1920

</div>

In this chapter we deal with the clustering of time series into meaningful groups, each of them identified through a probabilistic model representing the dynamic behavior of all the members of the group. This is achieved through model-based clustering, allowing the identification and description of several dynamic patterns of behaviors within a population. The patterns of behavior are characterized in terms of a probabilistic model.

Mixtures of distributions provide this approach with a modelling concept, which simultaneously describe the presence of several subclasses in a population, together with modelling kernels describing each of them. Mixture models have been employed since the end of the 19th century, with the papers of Newcomb [New86] from 1886 and the one of Pearson [Pea94] from 1894. The book of McLachlan and Peel [MP00] introduces this topic.

Many attempts to solve the problem of fitting a mixture of distributions in an efficient way were successful for specific cases [Has69, CM73]. In 1977, Dempster et al. [DLR77] described the *expectation maximization* (EM) algorithm for incomplete data. Several applications were described, amongst them, the fitting of mixtures models. Its simplicity, adaptability and general good performance makes it one of the most used inference techniques in presence of incomplete data. Particular cases are missing data [Har58, CBM71], finite mixtures models [Has69, CM73] and hyper-parameter estimation.

Mixtures of GLM for the modelling of multivariate data (although not for time series) is introduced in [WD95]. The mentioned model has been used on functional neuroimaging data [PF03] to obtain clusters of voxels characterized by a single time series, allowing the discovery of task-related connectivity between different parts of the brain.

Motivated by the latter mentioned methods from [FSK08] and [WD95], we develop an Expectation Maximization algorithm, which is adapted for the simultaneous clustering and fitting of mixtures of GLM for **multivariate long-term time series**.

Since in the scope of this thesis theoretical results are not proven, a simulation study is developed in Appendix A to test the different features of the proposed method.

5.1   Finite mixture of distributions

Following the approach given in [MMR05], a *finite mixture of distributions* is any convex combination

$$g(Y) = \sum_{h=1}^{H} p_h f_h(Y), \ \sum_{h=1}^{H} p_h = 1. \tag{5.1}$$

In many cases, f_h belongs to a parametric family of distributions, so that $f_h(x) = f(x, \lambda_h)$, where λ_h is the set of parameters for the $h-th$ component of the mixture. A common example is the mixture of Gaussians, parameterized by the mean μ_h and variance σ_h^2. From now on, the parametric finite mixture will be identified through $\{\theta_h\}_{h=1}^{H} = \{\pi_h, \lambda_h\}_{h=1}^{H}$.

The likelihood of a sample $\{y_i\}_{i=1}^{N}$ given the mixture $\{p_h, \lambda_h\}$ is

$$L(\{y_i\}_{i=1}^{N}) = \prod_{i=1}^{N} \sum_{h=1}^{H} p_h f(y_i, \lambda_h). \tag{5.2}$$

Is not difficult to see that the expansion of the inner sums in 5.2 leads to H^N terms, which already for $H = 2$ mixtures and $N = 10$ samples is 1024 terms in its expansion. This complexity makes the development of analytical solutions through maximum likelihood or Bayes estimators [MMR05] unfeasible.

5.2   Estimation maximization

In the following, the approach given by [Del02] is used to describe the basics of the estimation maximization algorithm for incomplete data.

Given observed data Y coming from a distribution parameterized by θ, the computation of a value θ^* maximizing the log-likelihood $L(Y, \theta)$ is desired. An analytical maximum likelihood estimator in presence of incomplete data is difficult to achieve, as shown for the specific case

of mixture models, due to the exponential amount of terms in the likelihood function. Instead, EM provides an iterative estimator which improves at every iteration the log-likelihood, being at the same time much simpler to compute.

Given observed data Y and some hidden variables $Z \in \mathcal{J}$, the estimation maximization algorithm pursues the estimation of the model parameter θ maximizing its posterior probability given Y, marginalizing over Z:

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \, log f(Y, \theta) = \underset{\theta}{\operatorname{argmax}} \, log \sum_{Z \in \mathcal{J}} f(Y, Z, \theta). \tag{5.3}$$

The log of the sum in Expression (5.3) makes its maximization very difficult. In the search for a more convenient expression, the above is rewritten as:

$$log f(Y, \theta) = log \sum_{Z \in \mathcal{J}} f^k(Z) \frac{f(Y, Z, \theta)}{f^k(Z)}, \tag{5.4}$$

where $f^k(Z)$ is a probability distribution over the space $\mathcal{J}$ of hidden variables, such that $\sum_{Z \in \mathcal{J}} f^k(Z) = 1$. Since the *log* function is concave, the Jensen's inequality states:

$$log \sum_{Z \in \mathcal{J}} f^k(Z) \frac{f(Y, Z, \theta)}{f^k(Z)} \geq \sum_{Z \in \mathcal{J}} f^k(Z) log \left(\frac{f(Y, Z, \theta)}{f^k(Z)} \right). \tag{5.5}$$

Suppose a starting approximation of θ, θ_k is known, the function

$$B(\theta, \theta^k) := \sum_{Z \in \mathcal{J}} f^k(Z) log \left(\frac{f(Y, Z, \theta)}{f^k(Z)} \right)$$

is defined, which fulfills $B(\theta, \theta^k) \leq log(f(Y, \theta))$, thus being a lower bound of the log-likelihood function. $f^k(Z)$ is an arbitrary distribution defined over $\mathcal{J}$ depending on the θ^k. From all the lower bounds given by the different possible distributions $f^k(Z)$, the optimal one in the point θ^k is chosen:

$$f^{k*} = \underset{f^k}{\operatorname{argmax}} \, B(\theta^k, \theta^k). \tag{5.6}$$

Using a Lagrange multiplier for the constraint $\sum_{Z \in \mathcal{J}} f^k(Z) = 1$, the objective function from (5.6) becomes:

$$F(f^k) = \lambda \left(1 - \sum_{Z \in \mathcal{J}} f^k(Z) \right) + B(\theta^k, \theta^k).$$

Applying the necessary condition for extrema with respect to $f^k(Z)$:

$$\begin{aligned}
\frac{\partial F}{\partial f^k(Z)} &= -\lambda + \frac{\partial B(\theta^k, \theta^k)}{\partial f^k(Z)} \\
&= -\lambda + \frac{\partial}{\partial f^k(Z)} \sum_{Z' \in \mathcal{J}} f^k(Z') log f(Y, Z', \theta^k) - f^k(Z') log(f^k(Z')) \\
&= -\lambda + log f(Y, Z, \theta^k) - 1 - log(f^k(Z)),
\end{aligned}$$

which yields:

$$
\begin{aligned}
f^k(Z) &= exp(-1 - \lambda)f(Y, Z, \theta^k) \\
&= \frac{f(Y, Z, \theta^k)}{\sum_{Z' \in (J)} f(Y, Z', \theta^k)} \\
&= f(Z|Y, \theta^k).
\end{aligned}
\tag{5.7}
$$

The found optimal lower bound has indeed the property of touching the likelihood function at θ^k since:

$$
\begin{aligned}
B^*(\theta^k, \theta^k) &= \sum_{Z \in \mathcal{J}} f(Z|Y, \theta^k) log \left(\frac{f(Y, Z, \theta^k)}{f(Z|Y, \theta^k)} \right) \\
&= \sum_{Z \in \mathcal{J}} f(Z|Y, \theta^k) log f(Y, \theta^k) \\
&= log f(Y, \theta^k) \sum_{Z \in \mathcal{J}} f(Z|Y, \theta^k) \\
&= log f(Y, \theta^k).
\end{aligned}
$$

$B^*(\theta, \theta^k)$ has a much more comfortable expression to maximize than the log-likelihood since it barely contains a sum of logs. Actually, it can be decomposed into:

$$
\begin{aligned}
B^*(\theta, \theta^k) &= \mathbb{E}_{log f(Z|Y, \theta^k)} f(Y, Z, \theta) - \mathbb{E}_{f(Z|Y, \theta^k)} log f(Z|Y, \theta^k) \\
&= \mathbb{E}_{f(Z|Y, \theta^k)} log f(Y, Z|\theta) + log f(\theta) - \mathbb{E}_{f(Z|Y, \theta^k)} log f(Z|Y, \theta^k).
\end{aligned}
\tag{5.8}
$$

Defining the expected complete log-likelihood

$$
Q^k(\theta) = \mathbb{E}_{f(Z|Y, \theta^k)} log f(Y, Z|\theta),
\tag{5.9}
$$

the prior on the parameters $f(\theta)$, and noting that $-\mathbb{E}_{f(Z|Y, \theta^k)} log f(Z|Y, \theta^k)$ (also known as the entropy of the distribution $f(Z|Y, \theta^k)$) does not depend on θ, one can proceed to the maximization of Expression (5.8) with respect to θ to obtain a new estimation θ^{k+1}, i.e.

$$
\theta^{k+1} = \underset{\theta}{argmax}\, B^*(\theta, \theta^k) = \underset{\theta}{argmax}\, Q^k(\theta) + log(f(\theta)).
\tag{5.10}
$$

The EM method can be finally summarized in the two steps:

- *E-Step*: Obtain the optimal lower bound in θ^k, $B^*(\theta, \theta^k)$ given by (5.7).

- *M-Step*: Find new estimates θ^{k+1} maximizing the optimal lower bound $B^*(\theta, \theta^k)$, as given in (5.10).

5.3   EM for finite mixtures

In the following, the E and M steps of the EM algorithm are derived for the special case of finite mixtures. For this the following elements need to be defined:

- hidden variables $Z_i \in \{0; 1\}^H$, s.t. $\sum_{h=1}^{H} Z_{ih} = 1$

- distributional assumptions on Z with respect to the model parameters, i.e. $f(Z_i|\pi, \lambda)$

- expression for the optimal lower bound $B^*(\theta, \theta^k)$

- maximization

5.3.1 Hidden variables

In the context of finite mixtures, a population is assumed to be decomposable into H subpopulations, and each individual of the population is assumed to belong to one of them. The unknown random vectors $Z_i \in \{0,1\}^H$ describe to which subpopulation an individual belongs, i.e. :

$$Z_{ih} = \begin{cases} 1, & \text{if individual } i \text{ belongs to subpopulation } h \\ 0, & \text{otherwise.} \end{cases} \tag{5.11}$$

The Z_i are assumed to be i.i.d multinomial with parameters $\{\pi_1, \ldots, \pi_H\}$ and $n = 1$, i.e.

$$f(Z_i|\pi, n = 1) = \prod_{h=1}^{H} (\pi_h)^{z_{ih}},$$

with $\pi = \{\pi_h\}_{h=1}^{H}$ being the mixing proportions for the mixture (5.1).

It is assumed furthermore that the observations $\{Y_i\}_{i=1}^{N}$ are conditionally independent given the covariates $\{Z_i\}_{i=1}^{N}$. The density of Y_i conditional on the Z_i under this assumption is defined as:

$$f(y_i|z_i, \theta) = \prod_{h=1}^{H} f(y_i|\lambda_h)^{z_{ih}},$$

and the complete log-likelihood

$$log f(y, z|\theta) = log f(y|z, \theta) + log(z|\theta) = \sum_{i=1}^{N} \sum_{h=1}^{H} z_{ih} log f(y_i|z_i, \lambda_h) + \sum_{h=1}^{H} z_{ih} log \pi_h. \tag{5.12}$$

5.3.2 E-step

In the E-step, given an estimation of the parameters $\theta^k = \{\lambda^k, \pi^k\}$, the optimal lower bound

$$\begin{aligned} B^*(\theta, \theta^k) &= \mathbb{E}_{Z|Y,\theta^k}[log f(Y, Z|\theta) - log f(\theta)] \\ &= \mathbb{E}_{Z|Y,\theta^k}[\sum_{i=1}^{N} \sum_{h=1}^{H} z_{ih} log f(y_i|z_i, \theta_h) + z_{ih} log \pi_h + log f(\theta)] \\ &= \sum_{i=1}^{N} \sum_{h=1}^{H} \mathbb{E}_{Z|Y,\theta^k}[Z_{ih}] log f(y_i|\theta_h^k) + log f(\theta) \end{aligned} \tag{5.13}$$

is obtained by computing the expectation $\mathbb{E}_{Z_{ih}|Y_i,\theta_h^k}[Z_{ih}]$. This can be achieved by means of the Bayes rule:

$$\begin{aligned} \mathbb{E}_{Z_{ih}|Y_i,\theta^k}[Z_{ih}] &= f(Z_{ih} = 1|Y_i, \theta^k) \\ &= \frac{f(Y_i,|Z_{ih} = 1, \theta^k) f(Z_{ih} = 1|\theta^k)}{f(Y_i|\theta^k)}, \end{aligned}$$

having

$$\begin{aligned} f(Y_i,|Z_{ih} = 1, \theta^k) &= f(y_i|\lambda_h^k), & \tag{5.14} \\ f(Z_{ih} = 1|\theta^k) &= \pi_h^k & \tag{5.15} \\ f(Y_i|\theta^k) &= \sum_{h=1}^{h} \pi_h^k f(Y_i|\lambda_h^k), & \tag{5.16} \end{aligned}$$

so that

$$\mathbb{E}_{Z_{ih}|Y_i,\theta^k}[Z_{ih}] = \frac{\pi_h^k f(Y_i,\lambda_h^k)}{\sum_{h=1}^h \pi_h^k f(Y_i|\lambda_h^k)}. \tag{5.17}$$

5.3.3   M-step

Let $a_{ih}^k = \mathbb{E}_{Z_{ih}|Y_i,\theta^k} Z_{ih}$. The objective function to be maximized in order to obtain a new estimation of the model parameters θ^{k+1} is:

$$B^*(\theta,\theta^k) \;\;=\;\; \sum_{i=1}^N \sum_{h=1}^H a_{ih}^k log f(y_i|\theta_h) + log f(\theta_h). \tag{5.18}$$

5.3.4   EM for finite mixtures of GLM

The last step of the methods employed for data analysis in this thesis includes the use of GLM as modelling kernels of the finite mixture, i.e.

$$f(Y_i,\lambda_h^k) = \sum_{t=1}^T log f(Y_{it},\beta_h^k,\omega_{it}^h,X_{it},...) = \frac{\theta' y_{it} - b(\theta_h^k)}{a(\lambda)}\omega_{it}^h + c(y,\lambda). \tag{5.19}$$

In the E-step, $a_{ih}^k = \mathbb{E}_{Z_{ih}|Y_i,\theta^k} Z_{ih}$ is computed maximizing (5.19) given estimations of $\{\beta_h^k\}_{h=1}^H$.

In the M-step, H GLM's are fitted through the ML estimator described in Chapter 4 with the weight $\omega_{it}^h = a_{ih} * \omega_{it}$.

5.4   Selecting the amount of clusters: BIC

One of the assumptions of the proposed methods is that the number of clusters is known. This means, a tool for the selection of the number of latent classes in a population has to be employed. Several techniques have been proposed in the literature for this aim. A short review of such techniques is reviewed in [FS11].

After analyzing several of the methods reviewed in the literature, we attached to the *Bayesian information criterion (BIC)*. This is a large sample version of the Bayes procedure based on the evaluation of the leading terms of its asymptotic expansion, thus not needing the knowledge of any priors. It is defined as the model-dimension penalized score:

$$BIC(y,\tilde{\beta}) = -2 * L(y,\tilde{\beta}) + Np * log(Nobs),$$

with $L(y,\tilde{\beta})$ the likelihood of the model given the maximum likelihood estimator $\tilde{\beta}$, Np is the amount of parameters, and $Nobs$ the number of observations used to fit the model. The selected model is the one minimizing the BIC. For $Nobs > 8$, BIC selects a more parsimonious model than its predecessor, the Akaike information criterion ($AIC = -2L + Np$).

In Appendix A it is shown that BIC performs very good in estimating the amount of clusters, when data is simulated from known finite mixtures with different amounts of groups (amount of

clusters, as well as all kernel parameters are initially known, which allows to evaluate how good BIC selects the real structure).

Chapter 6

Drinking patterns, evolution to and prediction of an alcohol addiction

Through the multivariate time series analysis framework presented in Chapters 4 and 5 the time series described in Chapter 3 are analyzed. This approach defines patterns of behavior that characterize the long-term drinking behavior of the tested Wistar rats under the *long-term alcohol self administration with repeated deprivation phases* protocol.

An evolution of the drinking behavior, as well as an acuteness of the loss of control symptoms in the alcohol drinking following a period of abstinence, can be inferred form the found dynamic patterns.

This chapter describes and interpretes the identified patterns, its correlations throughout phases, and finally how they relate to the ADE classification given in Chapter 3. The structure of the chapters can be summarized as follows:

- identification and characterization (modelling) of patterns of behavior during measured baseline and after-deprivation phases,

- characterization of long-time evolution in the drinking patterns,

- correlation of the found patterns of behavior during baseline and after deprivation phases and

- characterization and prediction of the ADE as a sign of alcoholism, given the observed patterns.

6.1   Statistical modelling and pattern selection

For the modelling of each time series it is assumed that the multivariate variable $Y_t = \{Y_t^r\}$ distributes multinomial with probability $\{\pi_t^r\}$ at each time point t and $r \in R$, where R is the set of presented solutions, i.e. $R = \{H_2O, EtOH5\%, EtOH10\%, EtOH20\%\}$. The probability vectors are modelled as a nonlinear function of a linear predictor $\eta_t = X_t\beta$ within the framework

of generalized linear models, since the multinomial distribution is a member of the FE[1].

All the individuals are assumed to have the same covariates at time t as described in Section 4.4.1. The probability of a drinking event of each solution is modelled through a different amount of covariates, selected through the procedure described in Section 4.4.2. At the beginning of the covariate selection, the covariate matrix for each solution $r \in R$ has the form

$$X_t^r = (x_{t,...}^r) = \left[1, \quad \{sin(\omega t) \ cos(\omega t)\}_{\omega \in \frac{2\pi}{T} * [1...20]} \right],$$

where T is the length of the time series. The original model has $4 * (41) = 164$ parameters per cluster. This is computationally very expensive to fit. After a covariates selection is made, only 65 components out of the 164 are found to significantly contribute to the model. For each solution $\{H_2O, EtOH5\%, EtOH10\%, EtOH20\%\}$ 37, 13, 10 and 5 covariates are selected respectively. This relates to the complexity of each solution time series, i.e. to the amount of times rats drink from each of the solutions (most from the water channel, and less often from the more concentrated alcoholic solutions).

Once the covariates are selected, the clustering procedure is performed for several amounts of clusters (H = 1, ... 10) and through the BIC, the "best" model is selected[2]. The selected model yields a classification of rats into patterns of behavior whose features are described in detail in Sections 6.2 and 6.3.

6.2 Baseline time series

6.2.1 Data

- Under the mentioned protocol, data for three different baseline phases (1st, 3rd and 5th) were recorded

- For each baseline phase time series were recorded during 4-5 weeks at 5 minute intervals.

- Measurements at time t are 4-dimensional vectors, with the amount drank at this time of each of 4 bottles: H_2O, $EtOH5\%$, $EtOH10\%$, $EtOH20\%$.

6.2.2 Preprocessing

Each baseline time series contains 20 days of 5 minute-wise measurements. Due to the day-night cycle (also called circadian rhythm or passive-active cycle), the presence of a periodic component is asserted with help of a Fourier transform of each of the solution time series.

Figure 6.1 shows a spike on the 20th Fourier component of each channel, signalling the presence of 20 cycles on the data, i.e. a day period. The periodicity of the time series allows a length reduction to only one day of measurements. The model is designed to describe a single characteristic day.

---

[1]see Chapter 4 for a description of the GLM framework

[2]best in the sense of describing better the data with the least amount of parameters possible
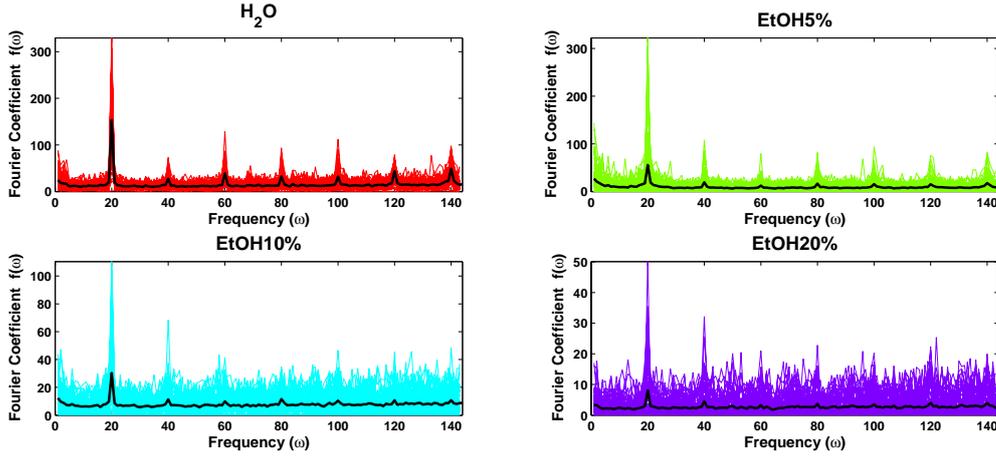
Figure 6.1: Fourier transform of each solution (color) and mean Fourier coefficients for all rats (black): data contains 20 cycles which are related to 20 days of measurements. The data can be seen as periodic with one day period (circadian rhythm).

Due to the complexity of the time series (multivariate, long-termed) the following simplification is made: instead of modelling the drinking amount, the probability of a drinking event of each solution at time t of the day is modelled. Drinking events are assumed to distribute multinomial with parameter n, and probability vector $\pi_t = \{\pi_t^r\}$, where n is the amount of measured days (in the case of the baseline data $n = 20$) and π_t^r is the probability of a drinking event at time t from solution $r \in R = \{H_2O, EtOH5\%, EtOH10\%, EtOH20\%\}$.

Let $\tilde{Y} = \{\tilde{Y}_t^r\}_{t \in T, r \in R}$ be the original 20 days-long time series of drinking amounts at time t from solution r. A new binary time series $\hat{Y}$ is obtained s.t.

$$\hat{Y}_t^r = \mathbb{1}_{\left[\tilde{Y}_t^r = 0\right]}.$$

The series $\hat{Y} = \{\hat{Y}_t^r\}$ states whether at time t a drinking event of solution r occurred or not.

Since the measurements are made every 5 minutes, one day long time series will have a length $T = 288$. The length of $\hat{Y}$ can be represented as $20 * T$. A new time series Y is defined so that

$$Y_t^r = \sum_{\substack{t' \in \{1 \dots 20 * T\} \\ t = t' \bmod T}} \hat{Y}_{t'}^r. \tag{6.1}$$

The new time series $Y = \{Y_t^r\}$ quantifies on how many days (out of the 20 measured) a drinking event of solution r occurred at time t. Y is a 20 times shorter series than $\hat{Y}$ but contains the same information, given the assumption of periodicity, i.e. π_t^r is the same every day. Under the mentioned assumptions $Y_t = \{Y_t^r : r \in R\}$ distributes multinomial with $n = 20$ and can thus be modelled within the GLM framework.

6.2.3   Statistical analysis

The best model for this data (selected through BIC) comprises 5 clusters. The time series related to $rat\_Id = 1$ are excluded from the analysis, since they are associated with an outlier like behavior (only this rat showed such a behavior).

Characteristics of the 5 drinking patterns

Cluster labels are given related to the mean daily net $EtOH$ intake, i.e. if $i < j$ then cluster i achieves on average a lower net alcoholic intake than cluster j. Figure 6.2 shows the mean daily net $EtOH$ intake of the five clusters.
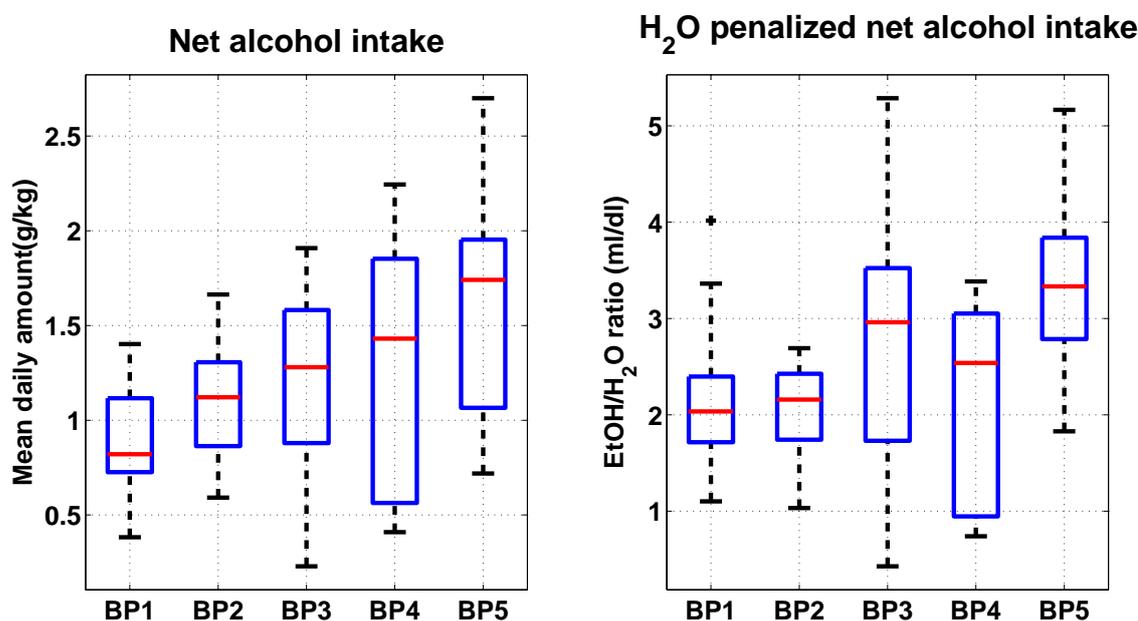

Figure 6.2: Box plot of daily net $EtOH$ consumption in g per body kg (left) and H_2O penalized net $EtOH$ intake in ml per H_2O dl (right) per pattern. The lower and upper lines represent the minimum and maximum intake within the group. The lower and upper limits of the box represent the 1st and 3rd quartiles.

Figure 6.3 (top) shows the probability of a drinking event for water, $EtOH5\%$, $EtOH10\%$ and $EtOH20\%$ for each of the five characteristic patterns at each time of the day. It can be seen how the day-night cycle is present: at 8 o'clock the lights are turned off (active cycle begins) and the rats start drinking actively with some spikes. At 20 o'clock, the lights are turned on and the passive period starts, accompanied by a decrease in the drinking activity. Figure 6.3 (bottom) shows a cumulative sum of the upper plotted probabilities, i.e. the expected amount of drinking events for each solution throughout the day per pattern.

Figure 6.4 shows the solution intake profile per pattern. Figures 6.5 and 6.6 show the preference for each solution per clusters from different perspectives. Figures B.3 and B.4 (see Appendix
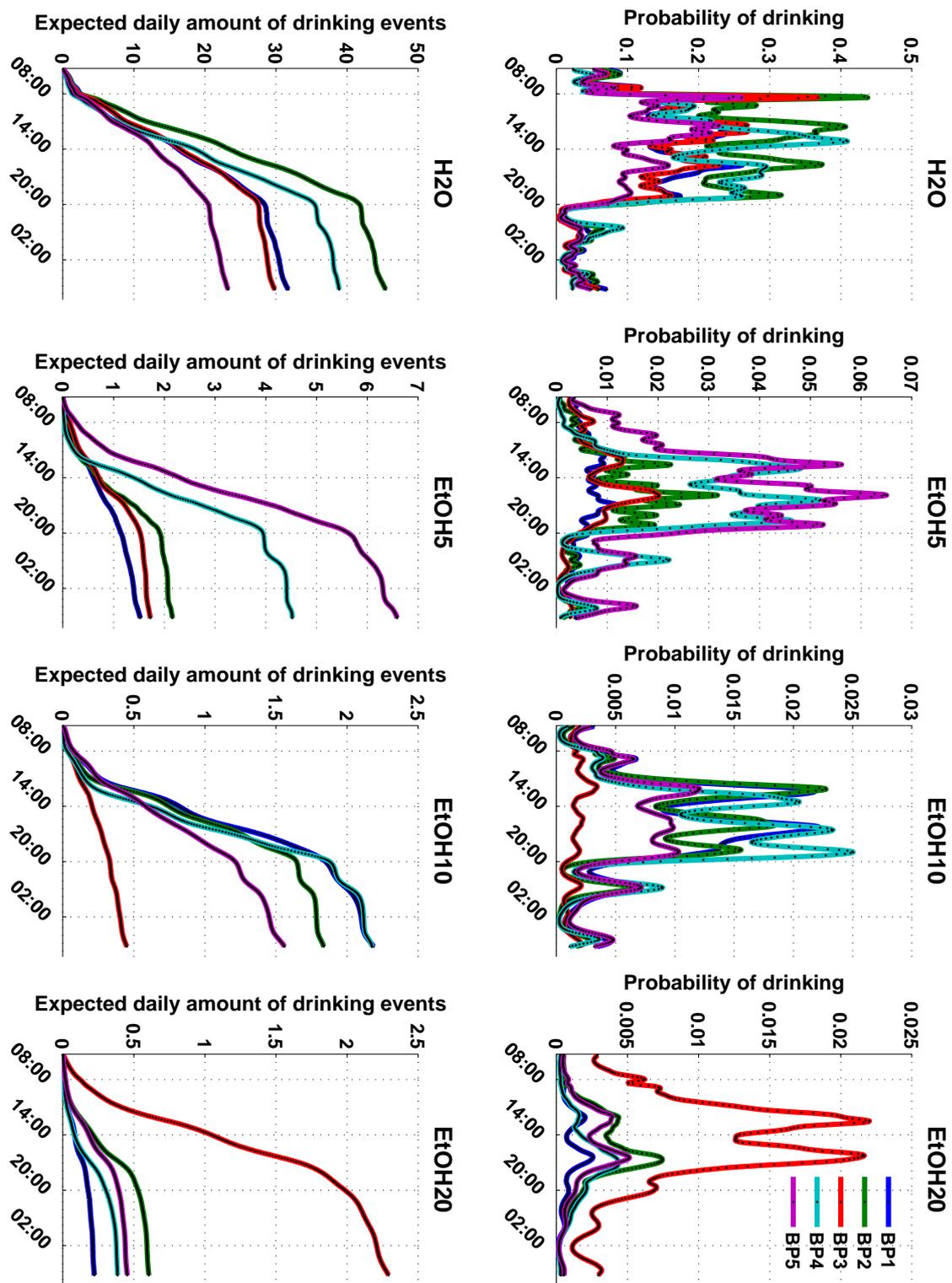
Figure 6.3: Characteristic patterns of behavior: Probability of a drinking event (top) and its cumulative sum interpreted as the expected amount of drinking events (bottom) at time t of the day from H_2O, $EtOH5\%$, $EtOH10\%$ and $EtOH20\%$ (left to right respectively)
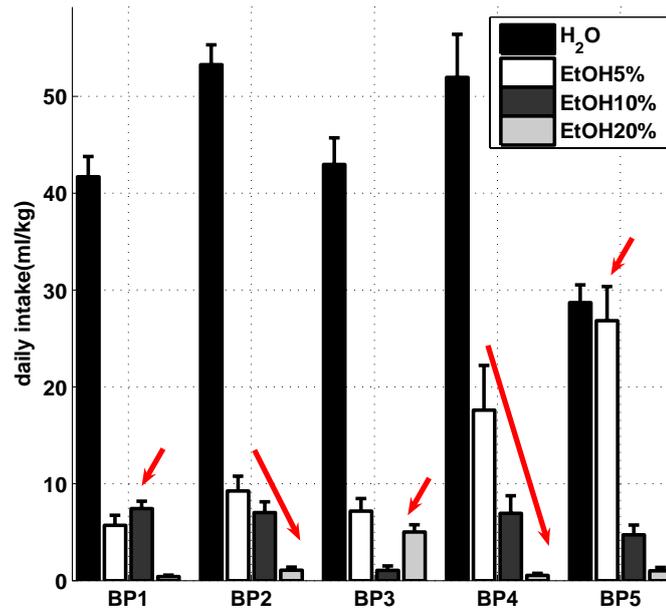
Figure 6.4: Solution intake profile per pattern

B) show the mean daily drinking amounts and the mean drinking amount per drinking event respectively per solution. They complete the description of the 5 characteristic clusters, which can be summarized as follows:

- BP1 achieves the *EtOH* intake mainly through 10% concentrated solution. From all the clusters the lowest net alcoholic intake (in terms of *g* per body *kg*) is achieved under this pattern. This can be related to human wine drinkers, who drink a glass of wine with each meal, but not much more.

- BP2 and BP4 are explorative drinking patterns: drinking amounts decrease with increasing concentration of the solution (beer is easy to drink, wine turns out to be a bit aversive, spirits taste way too strong). They are also characterized by a high water intake. The main difference between these 2 clusters is mainly given by the higher preference for *EtOH*10% of BP4, which is similar to the preference for *EtOH*5%.

- BP5 animals achieve the highest net *EtOH* mainly through 5% solution in huge amounts and some 20% concentrated solution. Very few water is drank throughout the day, and the final amounts can be compared with the consumed amounts of *EtOH*5%. In analogy to humans, this can be compared to beer drinkers, which often drink beer to quench the thirst. Animals under this pattern also achieve the highest net EtOH/H_2O ratio.

- BP3 is characterized by the preference for 20% concentrated solution. It has a huge variation regarding both *EtOH* representations.

- BP1, BP3 and BP5 can be considered as advanced patterns of behavior, characterized by a low frequency of H_2O drinking events and a preference for a determined solution concentration (10%, 20% and 5% respectively).

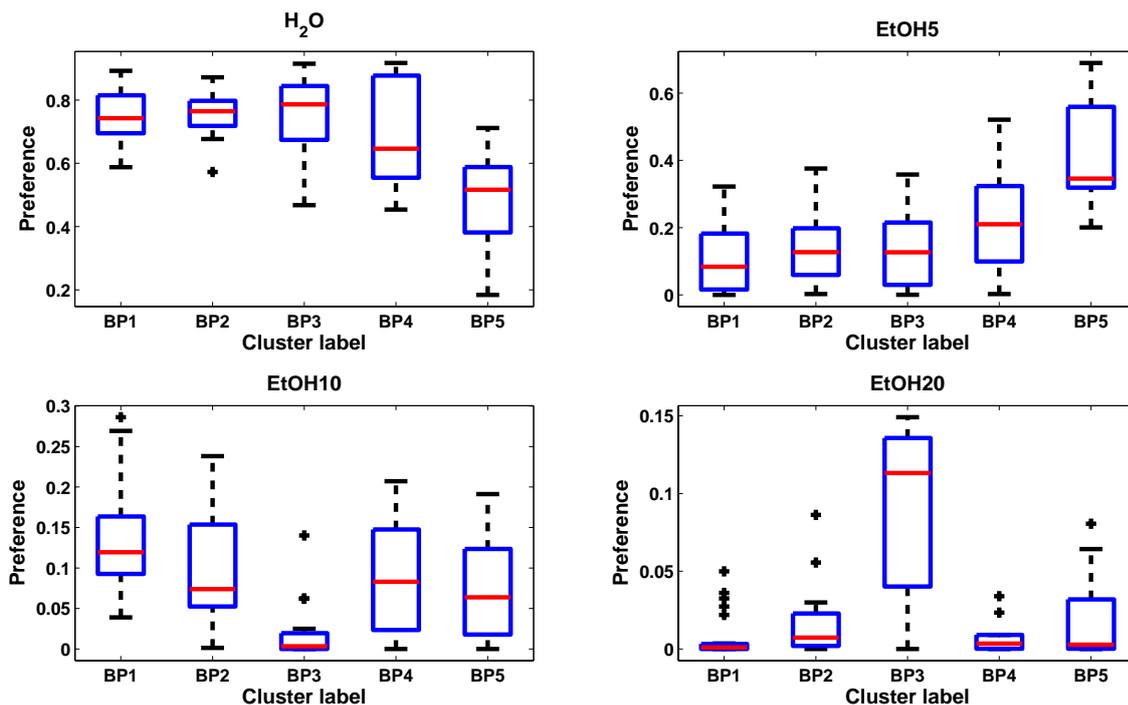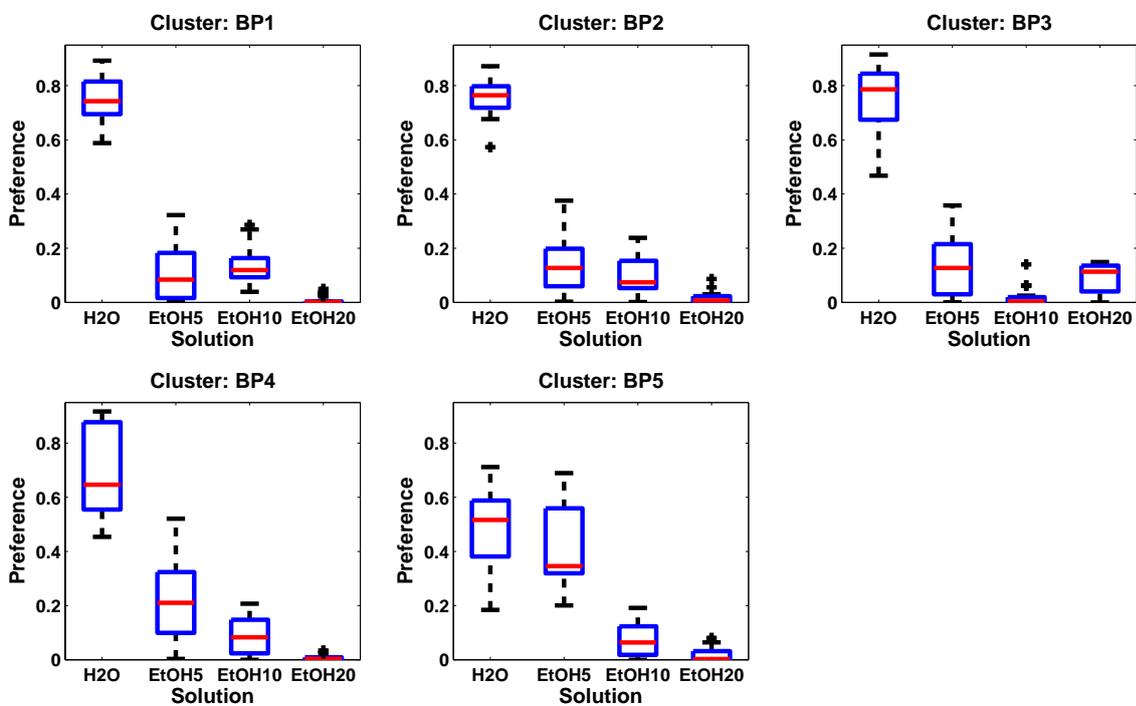Figure 6.5: Analysis of preference per cluster

Figure 6.6: Analysis of preference per solution

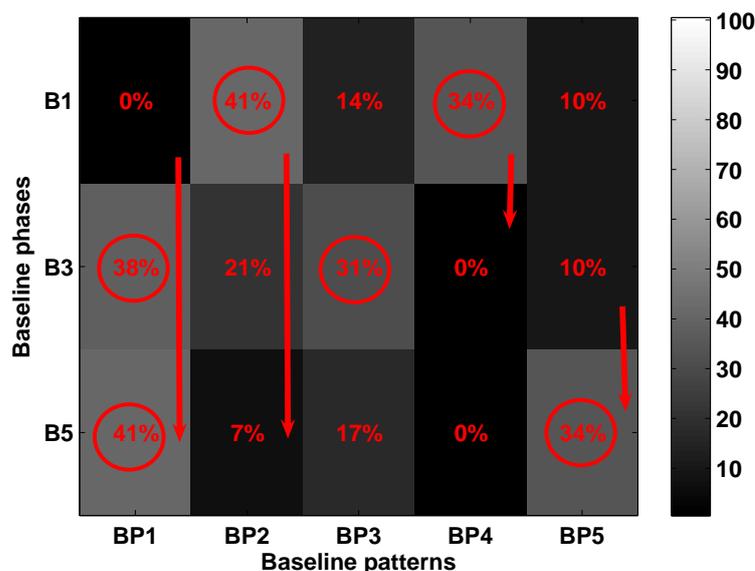**Distribution of patterns throughout baseline phases**

Figure 6.7: Distribution of patterns along the different baseline phases.

Once each pattern has been characterized, an analysis of the correlation between the different baseline phases can be made. Table 6.1 shows a summary of the distributional properties of the characteristic clusters for each baseline phase.

cluster label	mean net $EtOH$ (g)	baseline phase	preference
B1	0.495	3rd and 5th	$EtOH10\%$
B2	0.545	1st and 3rd	H_2O, $EtOH \uparrow \Rightarrow \text{Pref}\downarrow$
B3	0.580	3rd	$EtOH20\%$
B4	0.615	1st	H_2O, $EtOH5\%$, $EtOH10\%$
B5	0.890	5th	$EtOH5\%$

Table 6.1: Summary of the 5 clusters during the 3 measured baseline phases.

Figure 6.7 shows the distribution of the patterns for each baseline phase. Figure B.1 (see Appendix B) plots the classification of the rats on the different baseline phases. Our results allow to draw the following conclusions about the distribution of patterns along baseline phases:

- The first baseline phase is characterized by BP2 and BP4, both of them showing an explorative-like trend, since they have decreasing preference for increasingly concentrated solutions. BP4 disappears completely in the following baseline phases. BP2 tends also to extinguish in the 2 following phases. The main characteristic of this phase is the large water intake throughout the day.

- The third baseline phase is characterized by a great diversity of behaviors, where rats drink under the BP1, BP2, BP3 and BP5 patterns. It can be regarded as a "teenager phase

in the alcohol intake", where some rats still drink under one of the primary patterns of intake (BP2), some others have already developed a preference for a determined alcoholic concentration (BP5/5%, BP1/10% or BP3/20%).

- The fifth baseline phase is characterized by the BP1, BP3 and BP5 showing a strong preference for a specific concentrated solution.
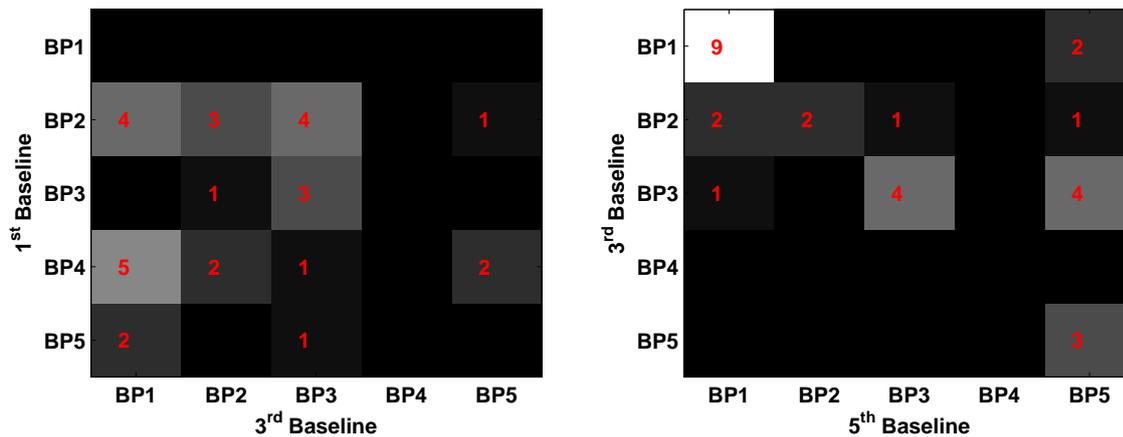
Evolution of the baseline drinking behavior

Figure 6.8: Left: pattern transition from 1st to 3rd baseline: unstructured, showing an instability in the drinking behavior (developing). Right: pattern transition from 3rd to 5th: more than half of the animals remain during the 5th baseline drinking under the same pattern as during the 3rd baseline.

An evolution of the drinking behavior throughout baseline phases can be inferred, where explorative-like patterns on the first phases are replaced by a solution-preference way of drinking in later stages. Figure 6.7 shows the distribution of patterns throughout phases where this transition can be observed. A more detailed view is given by Figure 6.8, where transitions from phase to phase are plotted.

- BP4 rats evolve from 1st to 3rd baseline into BP1,2,3,5 with 50% of them going to BP1. This is not surprising, since BP4 shows already a high preference for $EtOH10\%$, which is also characteristic of BP1. BP4 is not present in later baseline phases.

- Rats drinking under BP2 evolve from first to third baseline into BP1,2,3,5. There are only 2 BP1 rats on the fifth baseline; they showed BP1 throughout all the baselines, thus not evolving from their initial behavior.

- From the third to the fifth baseline phase, 18 out of the 29 rats (62%) remain stable in their patterns.

- Animals drinking under BP3 on the third baseline phase either remain or evolve to BP5.

6.3 after-deprivation (AD) time series

6.3.1 Data

- Data were recorded for three different after-deprivation (AD) phases (1st, 3rd and 5th) under the mentioned protocol.

- For each AD phase time series of 5 minute-wise measurements were recorded during the first 5 days after an alcohol deprivation phase.

- Measurements at time t are 4-dimensional vectors, with the amount drank at this time of each of 4 bottles : H_2O, $EtOH5\%$, $EtOH10\%$ and $EtOH20\%$.

- Rats were divided into two groups, holding 14 and 15 individuals respectively. The first group received quinine in the alcoholic solutions represented after the 1st deprivation phase. The second group received quinine after the 5th deprivation phase. The groups are denoted as Q (quinine) and C (controls), respectively, for the first AD phase, and C/Q respectively for the 5th AD phase. On the 3rd AD phase, no animal received quinine, being thus used as controls.

6.3.2 Statistical analysis

The AD time series are modelled again assuming a multinomial distribution with parameters $n = 1$ and probability vector $\pi_t = \{\pi_{tr}\}_{r \in R}$ in the GLM framework with seasonal components as covariates. The best model for this data set (selected through BIC) comprises $H = 3$ clusters, obtained through the described GLM based-EM algorithm.

Descriptive statistics of characteristic patterns

Descriptive statistics of characteristic patterns are performed separately for each Q and C group.

The three clusters are sorted according to the mean daily net $EtOH$ intake. Hence, the mean daily net intake for cluster i is greater than for cluster $j \iff i > j$.

The mean daily net $EtOH$ per cluster/group is shown in the box-plots of Figure 6.9. P-values from a t-test for 2 samples are shown in the color boxes. The net $EtOH$ intake of the C and Q groups is significantly different for patterns ADP2,3. However, the H_2O penalized $EtOH$ intake of both groups for patterns $ADP1$ and $ADP2$ is not significantly different at the 0.01 level.

The probability of a drinking event for a certain solution at each time t of the day for the three obtained clusters can be seen in Figure 6.11 (top). Its cumulative sum indicates the expected amount of drinking events throughout the day from each solution and is shown in Figure 6.11 (bottom).

The mean daily intake and the mean drinking amount per drinking event are shown in Figures B.5 and B.6 (see Appendix B).
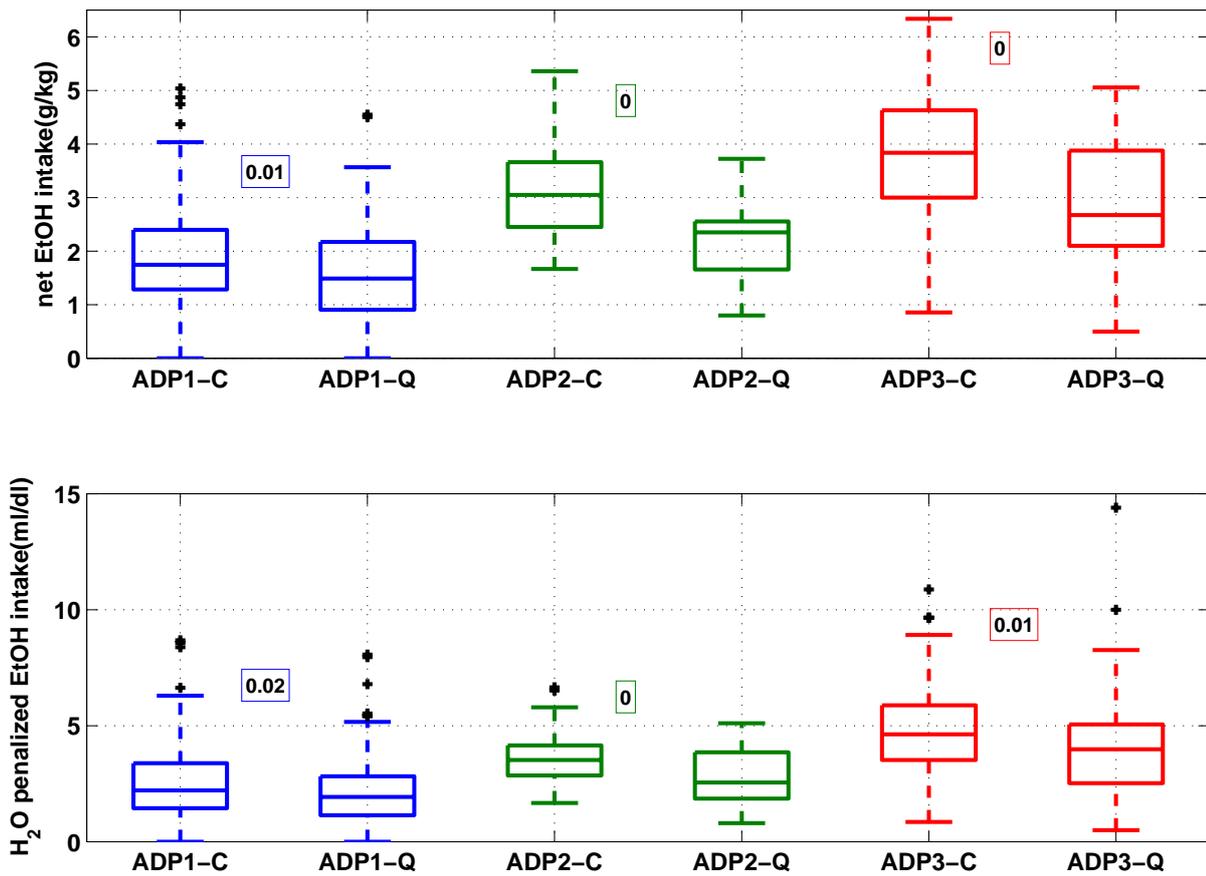
Figure 6.9: Mean daily net *EtOH* intake in *g* per body *kg* (top) and H_2O penalized *EtOH* intake in *ml* per *dl* of water (bottom) per cluster

Figure 6.10 shows the preference profile for each of the clusters, differentiating between control and quinine groups. Significant differences, tested through the two-sample *t*-test, are highlighted with colored boxes, specifying whether the alternative hypothesis is a decrease (red box, right tailed) or an increase (blue box, left tailed) of the intake.

With the aid of these figures we summarize the characteristics of the three clusters as follows:

- ADP1 is characterized by a very low preference for alcoholic solutions and a high water intake. Animals drinking under this pattern achieve the lowest daily net *EtOH* intake.

- ADP2 is mainly characterized by a very high preference for 5% concentrated solution. Animals drinking under this pattern achieve an average daily net *EtOH* intake.

- ADP3 is characterized by a high preference for alcoholic solutions and particularly for the 10% concentration. Animals drinking under this pattern achieve the highest daily net *EtOH* intake. Under this pattern, even during the passive period (before 8 o'clock), the probability of *EtOH* drinking events is higher than for the remaining patterns, and comparable to the probability during the active period.

- ADP1 quinine animals present increased water intake (left tailed 2 sample *t*-test p-val=0)
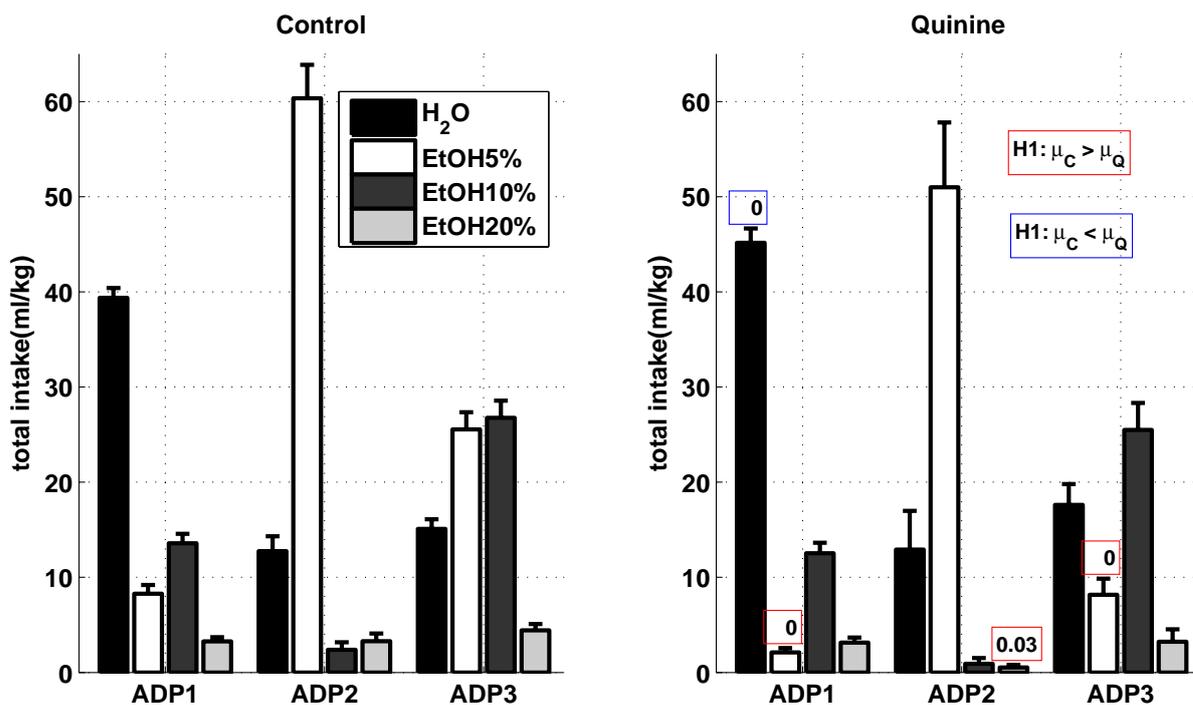
Figure 6.10: Solution intake profile of each after-deprivation pattern (ADP) per C/Q group. The red and blue boxes contain significant p-values obtained by means of a left/right respectively tailed 2 sample t-test, i.e. tests $H0 : \mu_C = \mu_Q$ vs $H1 : \mu_C > \mu_Q$ (right tailed, red box) or $H1 : \mu_C < \mu_Q$ (left tailed, blue box)

and a decrease in $EtOH5\%$ intake (right tailed 2 sample t-test p-val=0) in comparison with ADP1 controls.

- ADP2 quinine animals present a decrease in $EtOH20\%$ intake (right tailed 2 sample t-test p-val=0.03) in comparison with ADP2 controls.

- ADP3 quinine animals present a decrease in $EtOH5\%$ intake (right tailed 2 sample t-test p-val=0) in comparison with ADP3 controls.

Distribution and evolution of drinking patterns throughout deprivation phases

An analysis of the distribution of ADP for each AD day is made, comparing how it behaves along AD phases (see Figures 6.12 and 6.13). High p-values state that the null hypothesis of independence of the ADP distribution on the AD phase ($H0$) cannot be rejected, i.e. regardless of the AD phase, rats will present the same distribution of ADP. Low p-values state that $H0$ can be rejected, so depending on the AD phase, rats will present a different distribution of ADP. The following conclusions can be drawn:

- On the first AD day: control animals from the different AD phases presented the same distribution of ADP, namely ADP3 with high probability (p-value $= 0.06$, Figure 6.12 top-left); quinine animals presented a different distribution of ADP, according to the AD
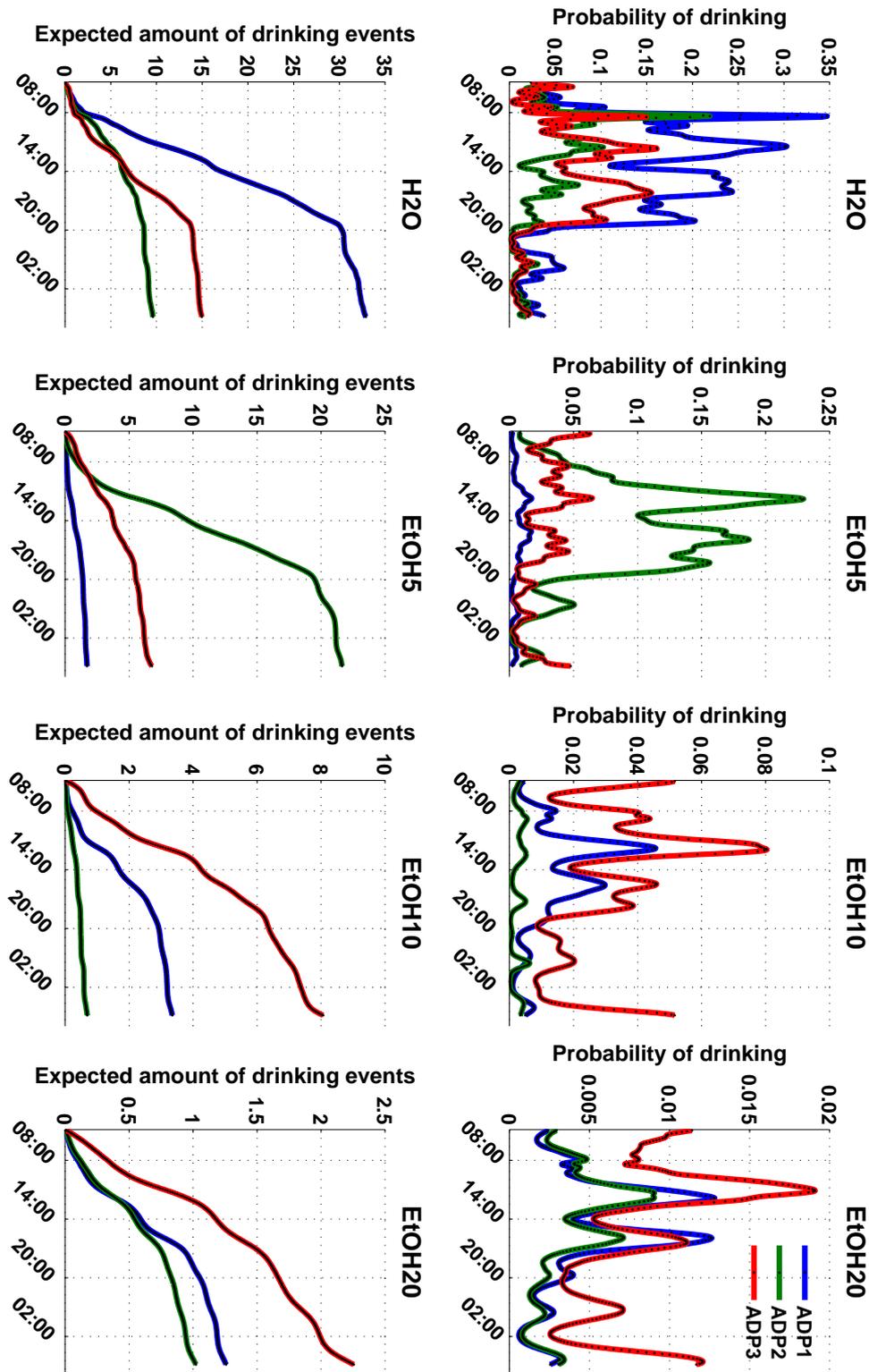
Figure 6.11: Probability of a drinking event at time t of the day (top) and its cumulative sum depicting the expected amount of drinking events until time point t of the day (bottom) of H_2O, $EtOH5\%$, $EtOH10\%$, $EtOH20\%$ (left to right).
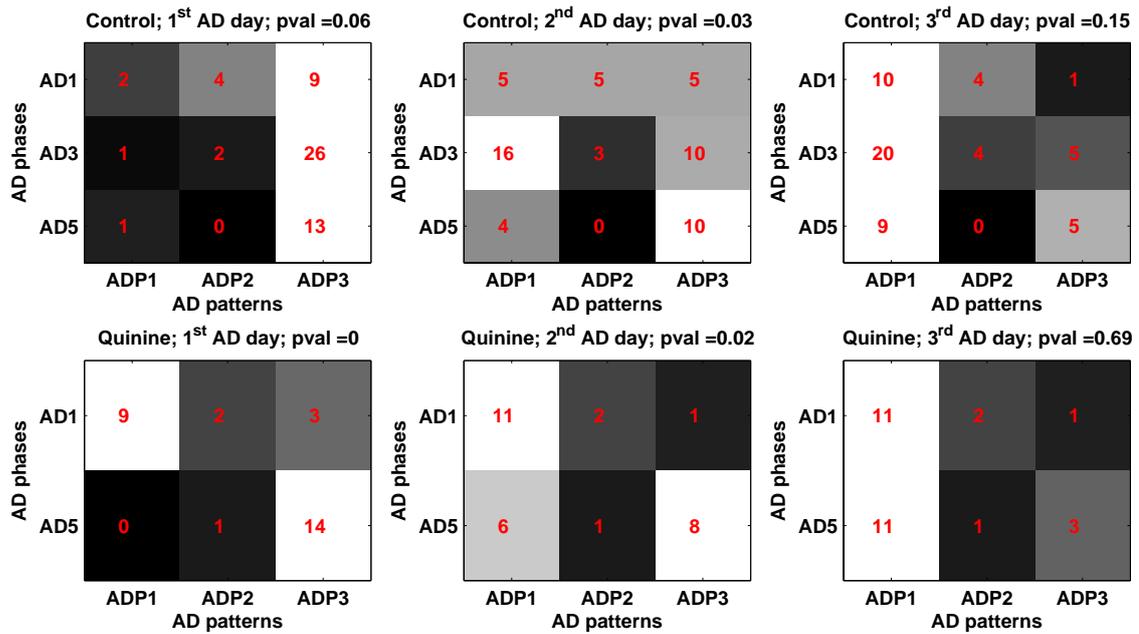
Figure 6.12: Distribution of ADP per AD phases for the control group (top) and the quinine group (bottom) throughout ADE days. The p-values are computed with the Fisher exact test for contingency tables to assert significant dependence in the distribution of ADP throughout phases per day.
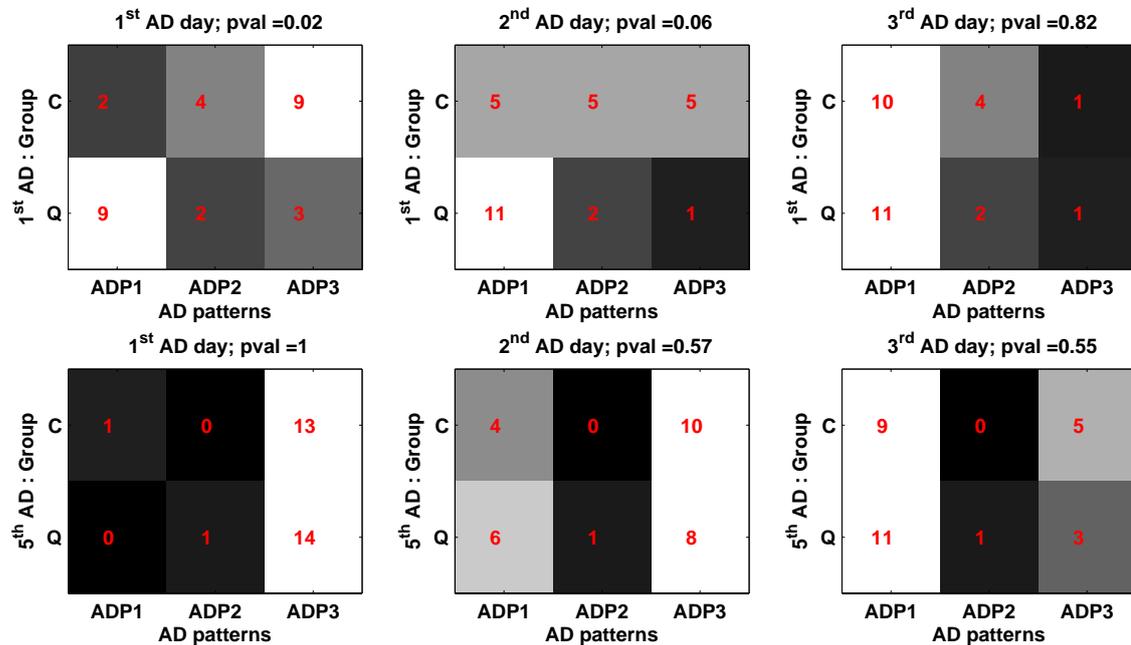
Figure 6.13: Same as Figure 6.12 but displayed differently to observe the inter-group differences for the 1st and 5th AD phases (top and bottom respectively). The p-values result from the Fisher exact test for contingency tables to assert significant dependence in the distribution of ADP throughout phases per day.

phase (p-value = 0, Figure 6.12 bottom-left). Comparing C/Q animals from the same AD phase (first column of Figure 6.13), one can see that there are differences between both groups on the first AD phase (p-value=0.02, Figure 6.13 top), but not on the fifth. These results state the development of inflexibility in the drinking at advanced AD phases.

- On the second AD day, a significant dependence of the distribution of ADP on the AD phase is observed. As general rules for both control and quinine groups:

  - ADP2 tends to disappear,
  - ADP3 tends to appear more frequently

 throughout phases (low p-values from Figure 6.12 second column). A separate analysis of both groups during each AD phase shows again, that Q/C groups differ during the first AD phase (p-value = 0.06, Figure 6.13 second column-top), while no significant difference is observed between Q/C during the fifth AD phase (p-value = 0.57, Figure 6.13 second column-bottom).

- On the third AD day, no significant dependence throughout phases for each group is observed (high $p-values$ from Figures 6.12 and 6.13, last column).

6.4 Relationship between BP and ADP

In order to find a correlation between patterns during baseline and those after a deprivation phase, transition matrices are obtained and shown in Figure 6.14. P-values are computed to establish significant differences between C/Q groups. Though one can see some differences, the very small sample size yields large p-values in all the cases, so no conclusions can be drawn regarding this issue.
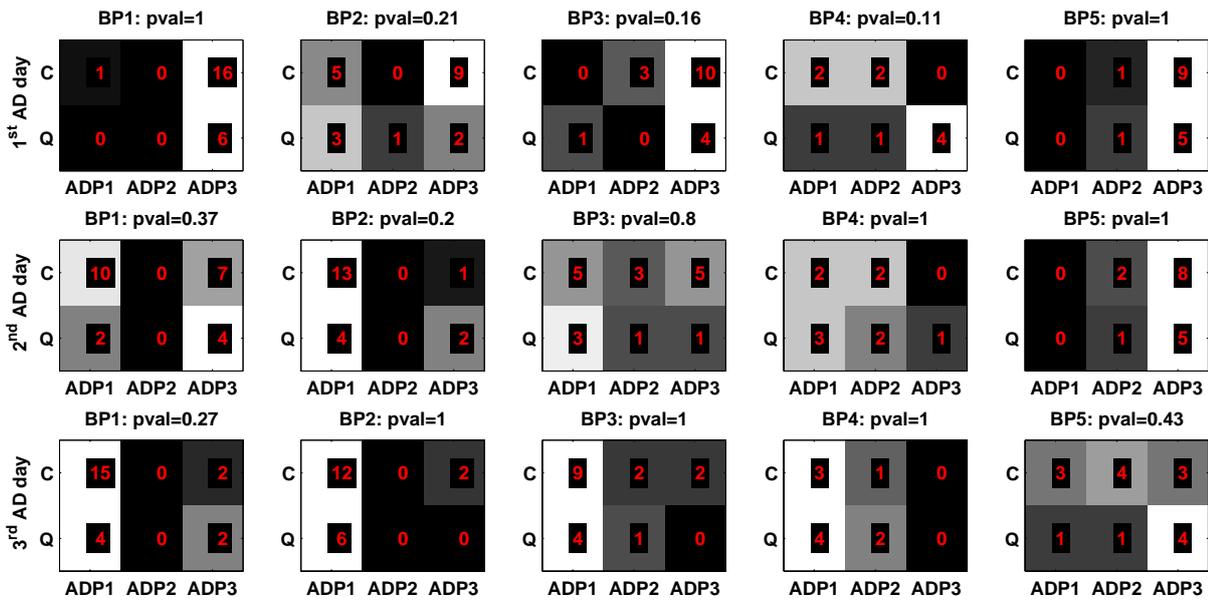

Figure 6.14: Transition matrices of baseline pattern to each after-deprivation day pattern.
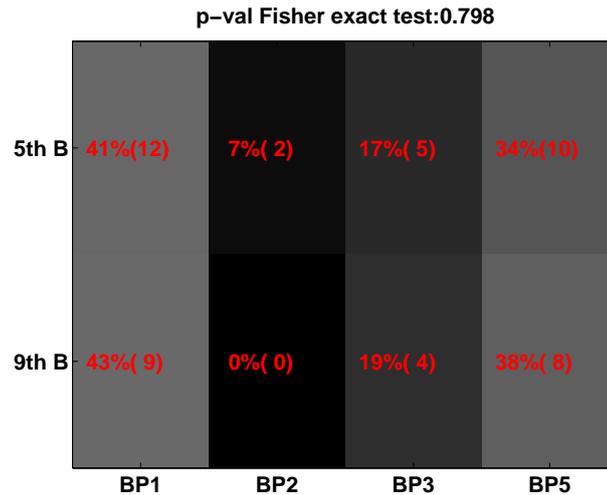
Figure 6.15: Distribution of BP during 5th and 9th baseline phases. The p-value is calculated with the Fisher exact test.

6.5   Robustness of BP and ADP

Baseline and after-deprivation patterns are found that dynamically develop throughout phases for a set of 29 rats. We hypothesized that after a certain period of time, rats have got to a stable point, where their behavior towards alcohol does not develop further. Under this assumption, the results obtained for the 5th cycle should also hold for more advanced cycles, given that these rats have been kept under the same conditions as the ones for which results were developed.

With this idea, time series of the 9th cycle of 22 male Wistar rats are included in the analysis. The baseline and after-deprivation time series are classified into the existing BP and ADP. This is achieved maximizing the likelihood of the series, given the fitted pattern parameters.

Figure 6.15 shows the distribution of BP during the 5th and 9th baseline phases. The p-value from the Fisher exact test states that no significant dependence on the baseline phase can be found, i.e. they have the same BP distribution. The only difference between the two cycles lies in BP2, which disappears on the 9th baseline. This corroborates the fact that BP2 is a primary pattern that tends to extinguish throughout deprivation phases.

Figure 6.16 shows the distribution of ADP during the 5th and 9th after-deprivation phases. The p-value from the Fisher exact test states that no significant dependence on the cycle can be found, i.e. they have the same ADP distribution. It can be seen that from the third day on, animals from the 9th ADE tend to remain in ADP3, corroborating the idea that the more AD phases, the longer this intense drinking pattern holds.

Figures 6.17 and 6.18 show the distribution of ADP given a certain BP of both phases: no significant dependence is observed. However due to the very small sample size, no strong conclusions can be drawn.

Figure 6.19 shows the intake profile for each BP (row-wise) for each group (column-wise), comparing both phases through a right/left tailed t-test (red/blue boxes depict significant decrease/increase of the intake on 9th phase with respect to the 5th). It can be seen that:

- BP1 rats ($EtOH10\%$ preferrers) decrease $EtOH5\%$ and increase $EtOH10\%$ intake on the 9th baseline with respect to the fifth, regardless of the group (C/Q). Control rats also reduce the water intake.

- BP5 control rats increase the $EtOH10\%$ intake.

- The rest of the BP/groups does not show significant differences.

As a general conclusion, the found BP patterns, as well as their preference profile holds from the fifth to the ninth baseline phases. The few behavioral changes between both phases are in the direction of an acuteness of the drinking behaviors (e.g. primary pattern BP2 completely disappears).

Figure 6.20 shows the intake profile for each ADP (row-wise) for each group (column-wise), comparing both phases through a right/left tailed t-test (red/blue boxes depict significant decrease/increase of the intake on 9th phase with respect to the 5th). It can be seen that:

- A reduction in the $EtOH5\%$ intake is observed in all ADP/groups in the 9th phase with respect to the fifth.

- ADP3 control animals increase the $EtOH10\%$ and decrease the $EtOH20\%$ intake. ADP3 quinine animals decrease the $EtOH10\%$ intake.

As a general conclusion, after the 9th deprivation phase, rats avoid the less concentrated solution. A preference for the $EtOH10\%$ solution prevails. The intake of $EtOH20\%$ shows no interesting features.

6.6 ADE analysis

The topic of this section is to deal with the final goal of this thesis: predicting the risk of an alcoholic addiction at early phases of alcohol intake. For this, the classification obtained through the described procedure in Chapter 3 is correlated to the characteristic BPs.

The inspection of the contingency tables between BPs and the ADE (see Figure 6.21) classification yields no significant dependencies. This has of course to do with two important facts: 1) the small size and 2) the fact that most of the Wistar rats are known to present ADE. In the presence of quinine, they might need a longer time (more deprivation phases) to present ADE. Already on the fifth AD phase, ADE regardless of quinine can be observed.

BP2 shows a statistically significant dependency on the C/Q group (p-value=0.03), stating a vulnerability to quinine, which is not surprising, since it is mostly related to the 1st AD phase.
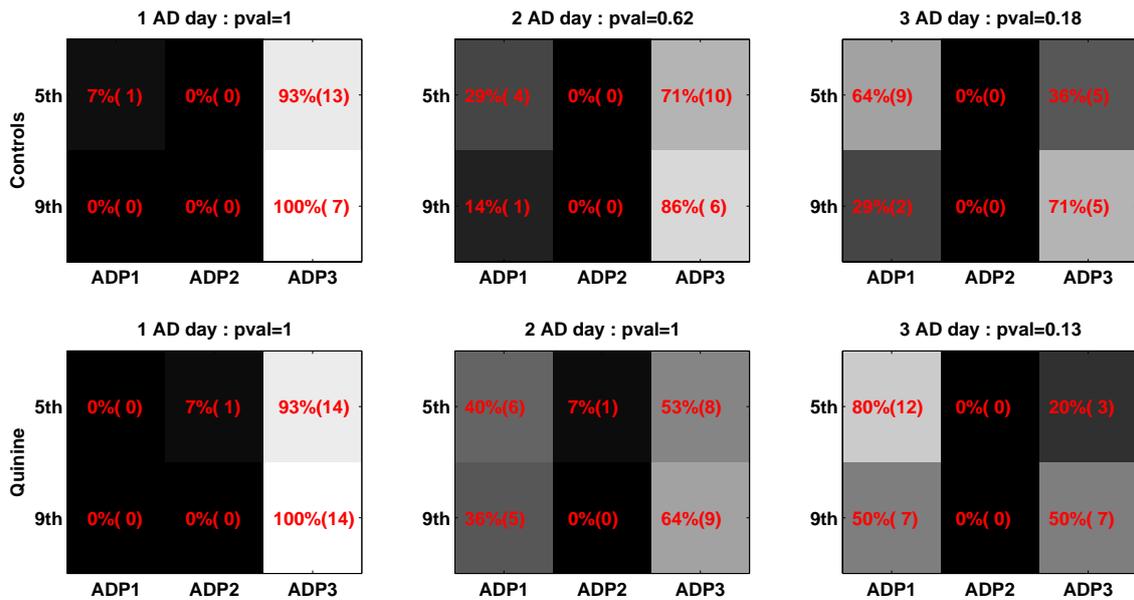
Figure 6.16: Distribution of ADP during the 5th and 9th after-deprivation phases. The p-value is calculated with the Fisher exact test.
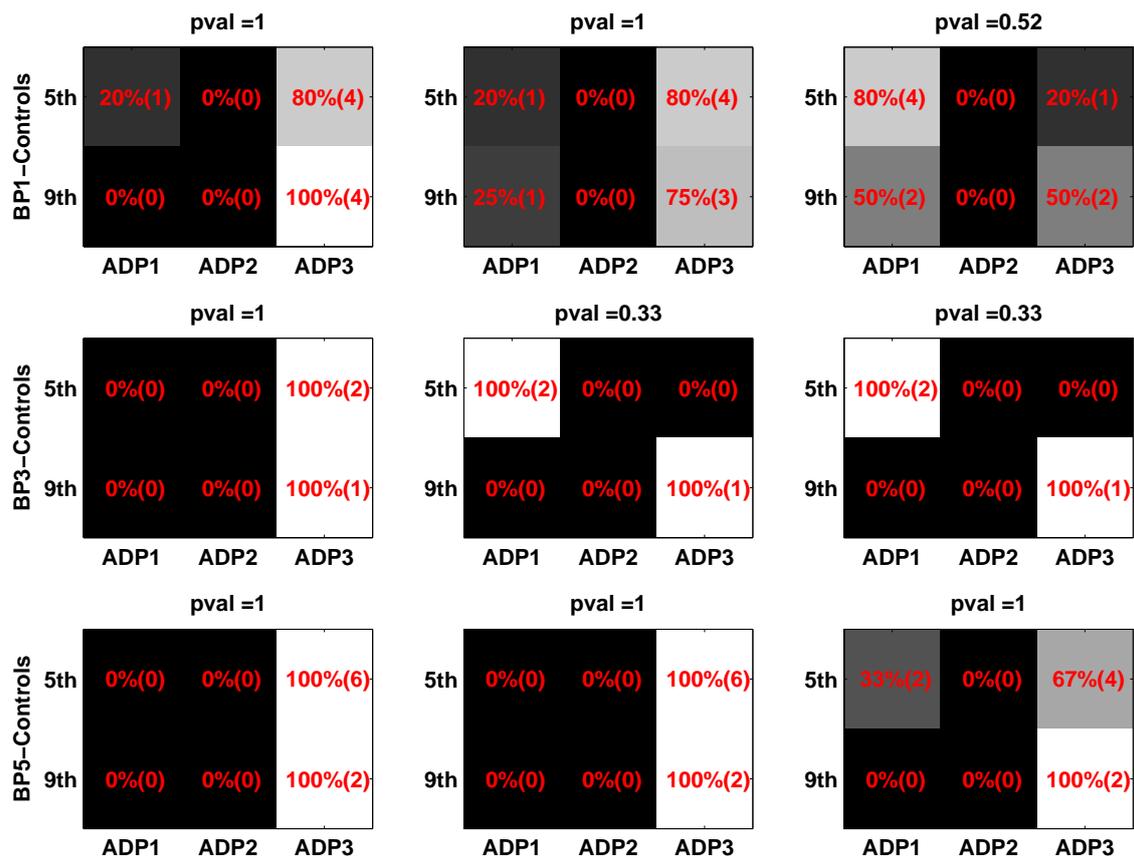
Figure 6.17: Distribution of ADP given BP during the 5th and 9th cycle for the control group. Only BP1,3,5 are analyzed. BP2 was presented only by two animals during the 5th baseline and no animal from the 9th baseline. The p-value is calculated with the Fisher exact test.
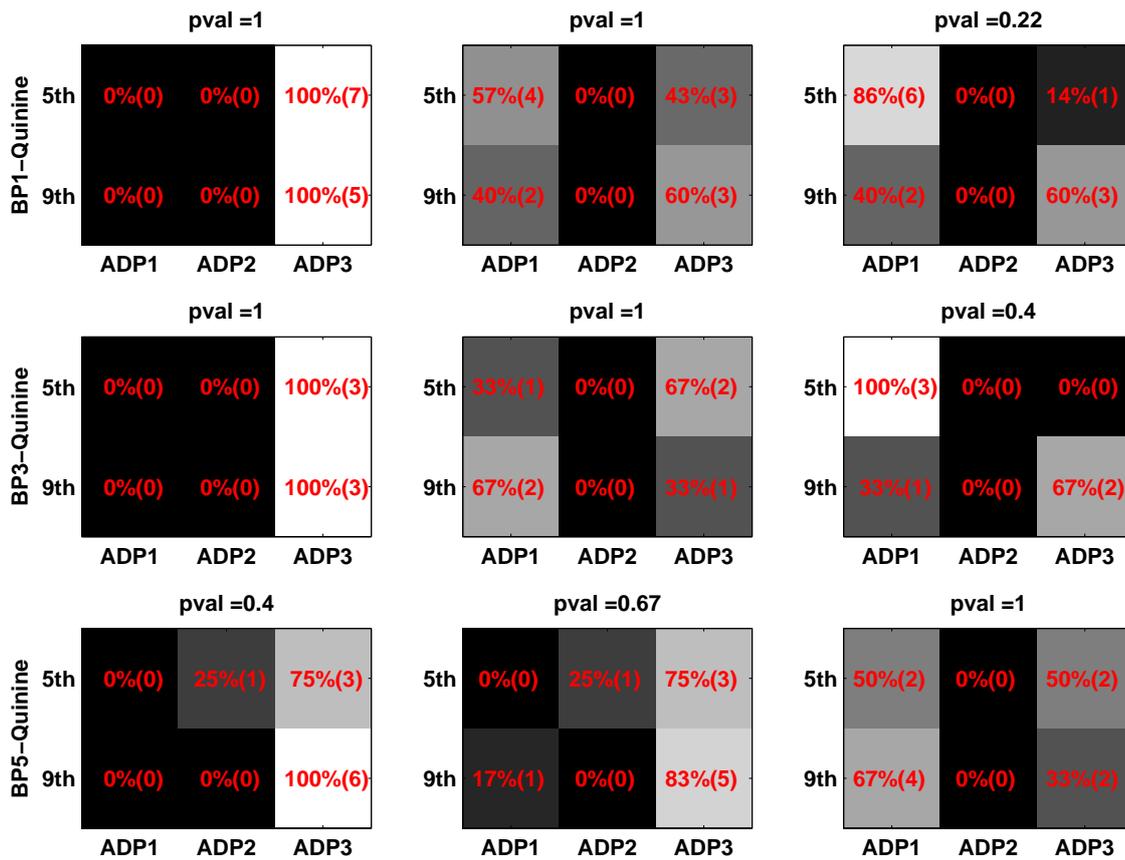
Figure 6.18: Distribution of ADP given BP during the 5th and 9th for the quinine group. Only BP1,3,5 are analyzed. BP2 was presented by only two animals during the 5th baseline and no animal from the 9th. The p-value is calculated with the Fisher exact test.
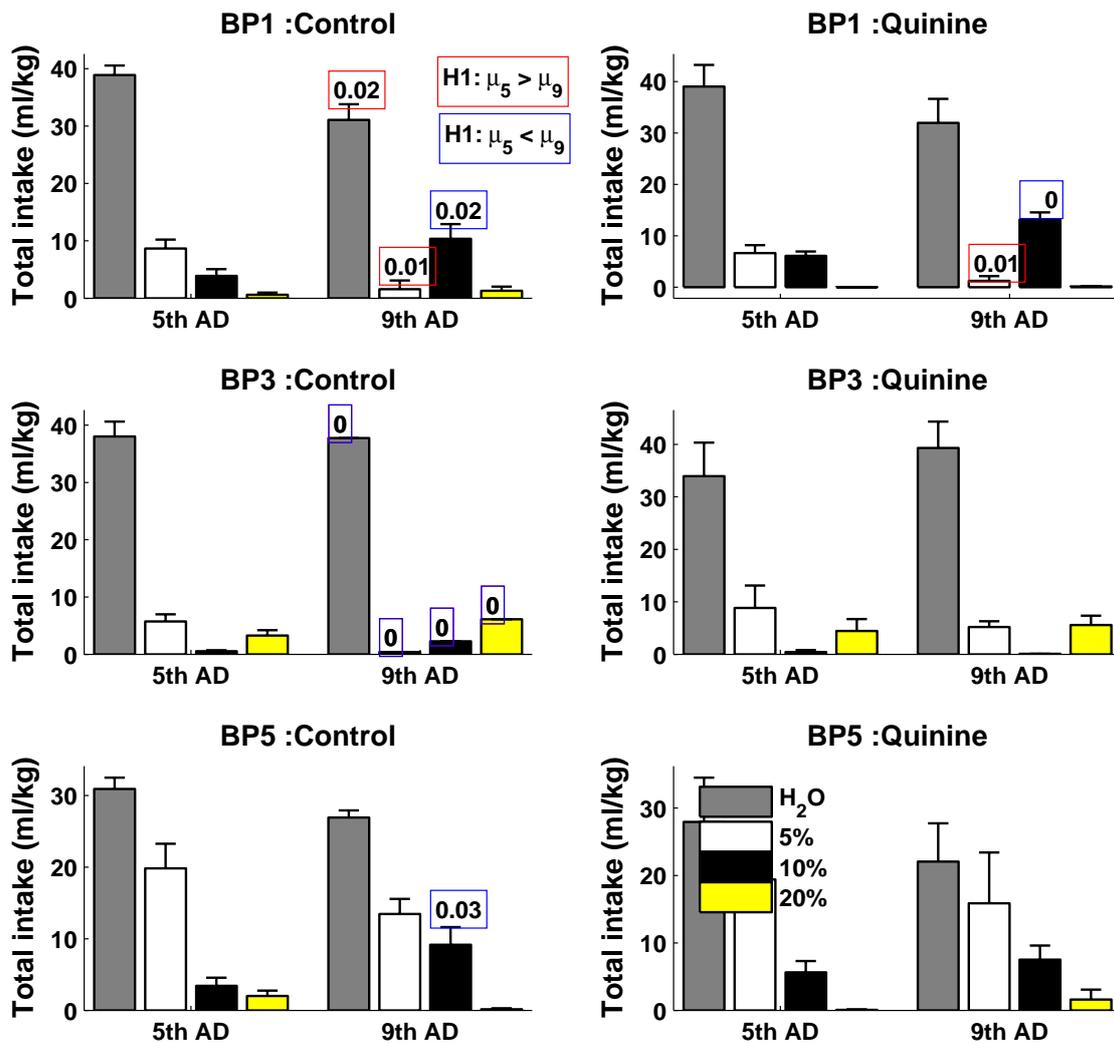
Figure 6.19: Intake profile per BP for each phase for controls (left) and quinine (right) animals.
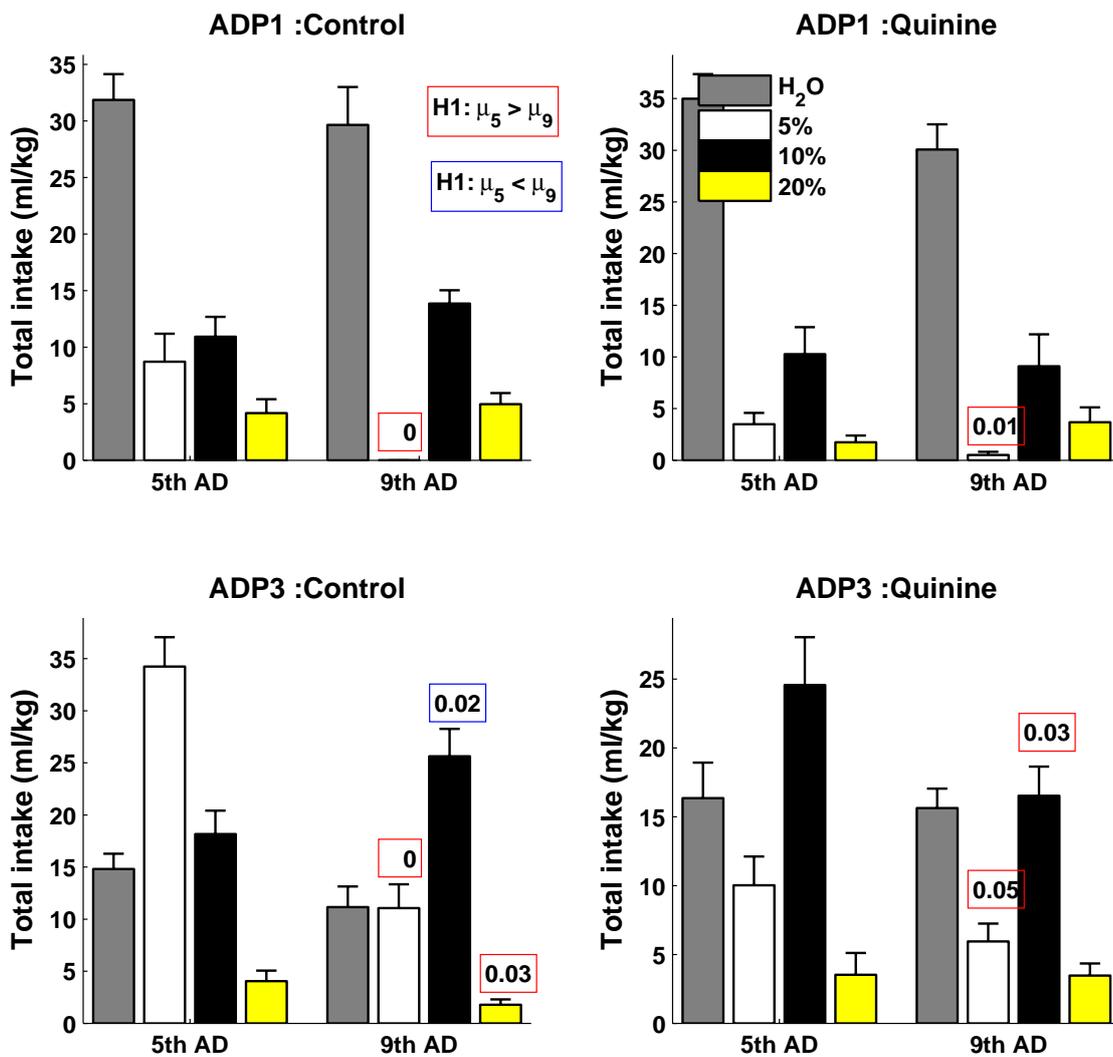
Figure 6.20: Intake profile per ADP for each phase for controls (left) and quinine (right) animals.
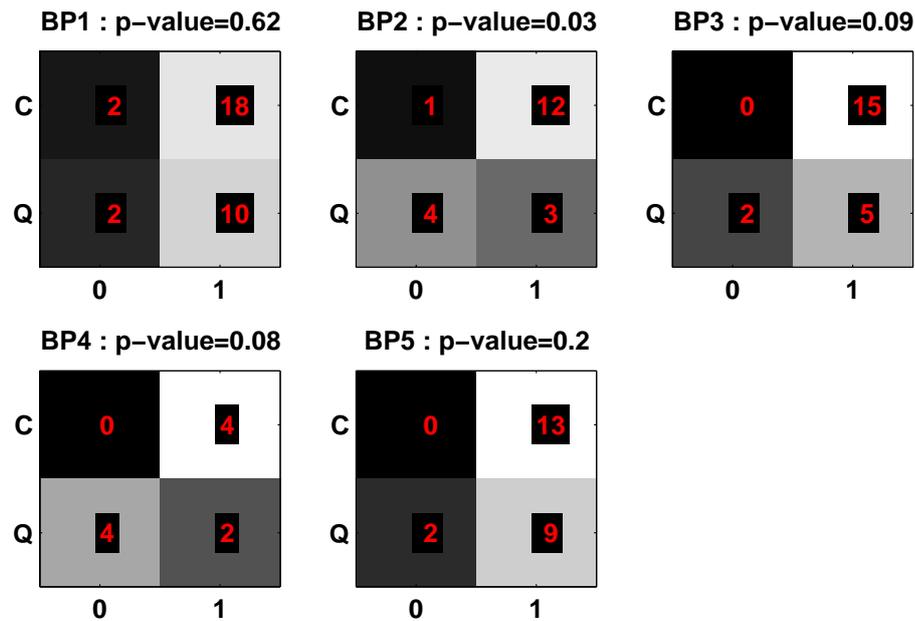
Figure 6.21: Relationship between Q/C groups and ADE classification for each BP

BP4 has a low p-value $= 0.08$, though greater than 0.05. Since this pattern contains the smallest sample size, no strong conclusions should be made based on this p-value.

BP1,3,5 have almost the same distribution of ADE $= 0/1$ for both C/Q groups (with p-values 0.62, 0.09 and 0.2 respectively).

Since most animals on the 5th AD present ADE independently from their pattern or whether they receive quinine or not, it makes no sense trying to predict from early phases of intake whether they will present this symptom of alcohol dependence or not. However, the described procedure could be used for groups of animals where the presence of an ADE is more heterogeneous. Here it would make sense to identify features that differentiate a group of animals that have developed an addiction from those who did not.

6.7 Summary of results and conclusions

In the analyzed baseline data five pattern of behavior could be identified:

- BP2 and BP4 can be regarded as initiating behaviors towards alcohol: $EtOH5\%$ tastes almost sweet, $EtOH10\%$ does not taste so bad, and $EtOH20\%$ has a very aversive taste. Thus beginner rats drink consequently the less the more concentrated the solution. They achieve a high water intake (by drinking very frequently). Most of the animals from the first baseline phase drink under these two patterns.

- Patterns BP1, BP3 and BP5 can be regarded as advanced, since they show already a preference for an alcoholic solution (10%, 20% and 5%). Most of the animals from the

third and fifth phases drink under these patterns.

- BP1 rats can be regarded as very moderate drinkers. They appear already on the 3rd baseline and most of them remain during the 5th.

- BP3 is mostly present during the 3rd baseline phase. This could signal a transition behavior in the development towards a final drinking style.

- BP5 rats achieve the highest net $EtOH$ consumption and are mostly characteristic of the 5th baseline.

In the analysis of the AD time series 3 clusters are identified:

- ADP1 is characterized by the lowest preference for alcoholic solutions and the highest water intake. Animals drinking under this pattern achieve the lowest daily net $EtOH$ intake.

- ADP2 is mainly characterized by a very high preference for 5% concentrated solution.

- ADP3 is characterized by a high preference for alcoholic solutions and particularly for the 10% concentration. Animals drinking under this pattern achieve the highest daily net $EtOH$ intake and drink at unusual times of the day. This pattern is mostly present on the first days after a deprivation phase.

- ADP1 and ADP3 quinine animals present a decrease in $EtOH5\%$ in comparison with their respective controls.

- ADP2 quinine animals present a decrease in $EtOH20\%$ intake in comparison with ADP2 controls. This pattern appears mostly during the 1st and 2nd AD phases for several days. It is rare during advanced AD phases.

- On the first AD day, the distribution of ADP of the quinine groups showed a significant dependence on the AD phase: during the first AD, most rats drank under ADP1; during the 5th AD most of the rats drank as controls did, under ADP3.

- On the second AD day, a significant dependence on the AD phase is observed. As general rules for both control and quinine groups:

  - ADP2 tends to disappear,
  - ADP3 tends to appear more frequently

  throughout phases.

- On the third AD day, no significant dependence throughout phases for each group is observed.

A comparative study is made between 5th and 9th cycles, where rats from the latter are different to the ones used for the pattern modelling. The following is concluded:

- The distribution of BP is independent of the baseline phase: in advanced phases (like the 5th and 9th) the probability of observing $BP1, 2, 3, 4, 5$ is $[0.42, 0.04, 0.18, 0, 0.36]$ respectively.

- The distribution of ADP is independent of the AD phase: It can be seen that from the 3rd day on, animals from the 9th ADE tend to remain on ADP3, corroborating the idea, that the more AD phases, the longer this intense drinking pattern holds.

Few changes are observed regarding the intake profile of each BP for each group of both phases:

- BP1 rats ($EtOH10\%$ preferrers) decrease $EtOH5\%$ and increase $EtOH10\%$ intake on the 9th baseline with respect to the fifth, regardless of the group (C/Q). Control rats also reduce the water intake.

- BP5 control rats increase the $EtOH10\%$ intake.

- The rest of the BP/groups does not show significant differences.

As a general conclusion, the found BP patterns, as well as their preference profile hold from the fifth to the ninth baseline phases, and the few observed changes are towards the acuteness of their drinking behaviors (less water, more alcohol).

Few changes are observed regarding the intake profile of each ADP for each group of both phases:

- A reduction on the $EtOH5\%$ intake is observed in all ADP/groups on the 9th phase with respect to the fifth.

- ADP3 control animals increase the $EtOH10\%$ and decrease the $EtOH20\%$. ADP3 quinine animals decrease the $EtOH10\%$ intake.

As a general conclusion, after the 9th deprivation phase, rats avoid the less concentrated solution. A preference for the $EtOH10\%$ solution prevails. The intake of $EtOH20\%$ shows no interesting features.

Regarding the prediction of ADE from baseline drinking, it can be seen that BP2 and BP4 rats are vulnerable to quinine taste affected solutions. BP1, BP3 and BP5 rats, tend to present ADE regardless from the taste of the alcoholic solutions.

The prediction of presenting ADE at advanced phases from early drinking stages is not interesting within the scope of this thesis. Most of the animals present ADE after several deprivation phases regardless of the taste adulteration. I.e. independently from their initiation in the alcohol drinking, they will present inflexibility in the drinking, thus ADE. So nothing needs to be predicted, since it is known that the outcome will be positive in almost all the rats.

More heterogeneous data sets could be analyzed, where the ADE is more meaningful (as explained, Wistar rats are known to present ADE after several deprivation phases even when quinine is added to the alcoholic solutions). Following the described procedure, preliminary

patterns of behavior could be identified, that could signal a future alcoholic dependence.

Though the original goal of prediction of ADE from early baseline patterns is not applicable in the scope of this thesis (due to the data limitations), the developed methodology remains. It allows the identification of early dynamic drinking patterns conditioning an addiction to alcohol, as well as the pathway of the evolution from initial patterns of alcohol drinking into advanced sick patterns.

Chapter 7

Discussion and conclusions

7.1 Summary of the study

In this thesis we have dealt with the classification of time series into dynamic patterns. The patterns have been described in terms of dynamic statistical models and can evolve throughout long time periods. The proposed methodology has been used to define alcohol drinking patterns of Wistar rats and to observe the pathway towards the development of an addiction to alcohol.

7.1.1 Methodology for identification and modelling of dynamic drinking patterns and their evolution

As a first step, we model patterns of behavior in terms of *generalized linear models (GLM)* (see Chapter 4). A generalization of the GLM framework to allow for panels of time series is described, which allows the modelling of several types of uni- and multi-variate panels of time series, as long as they are samples of a member of the family of exponential distributions. Such an approach was already addressed in different time series modelling contexts [FK87, Li94, Pru93, FK98]. Its flexibility for modelling several multivariate distributions, allowing the inclusion of time varying covariate information, makes this framework a good choice for the modelling kernels of interest.

In a second step, generalizing the results of [WD95], we developed an *estimation-maximization (EM)* algorithm (see Chapter 5), which fits a finite mixture of GLM to panels of uni- or multivariate time series. A procedure for selecting the "best" model through BIC is proposed and later on tested on simulations in Appendix A. The developed methods belong to time series model-based clustering approach, which has been addressed by several authors, by fitting mixtures of specific models, e.g. Markov Chains [FSK06] and binomial processes based on either logit or probit link functions to model probabilities of success [ZZ04, ABH11].

We allow with our generalization the classification of multivariate time series into a small amount of dynamical patterns. The latter are modelled in terms of GLM with time varying covariates.

In order to explore the performance of the algorithms on data for which our assumptions hold

(namely that conditioned on the covariates the observations are independent across time and individuals), a simulation study is presented (see Appendix A). Indication of the fulfilling of the nice theoretical properties of consistency and asymptotic normality are observed. The EM-classifier finds the underlying structure in most of the 100 simulations experiments for each number of clusters $h = 1; \ldots; 5$.

7.1.2   Application: behavioral studies on alcoholism in animals

The proposed methodology was applied to time series from behavioral studies on alcoholism, yielding a dynamic characterization and evolution of ethanol drinking patterns of Wistar rats under the *long term alcohol self administration with repeated deprivation phases protocol*. This is an experimental set up, aiming to induce high alcohol baseline intake, withdrawal symptoms during abstinence and relapse features after representation of alcohol. After deprivation of alcohol, rats present the *alcohol deprivation effect* and after several deprivation phases, they display inflexible drinking by increasing intake with respect to baseline levels, even if the alcoholic solutions are adulterated with quinine. This inflexibility is a clear sign of loss of control and compulsion in the alcohol drinking, thus a symptom of a dependency.

Under the described protocol, time series from the first 1st, 3rd and 5th baseline and after-deprivation phases are recorded. As a result, the drinking behavior of a rat, from early to advanced stages is finely time-wise described, i.e. in terms of drinking amounts of each solution every 5 minutes.

The raw inspection of the large amount of provided data by the naked eye is limited. It is difficult -or perhaps impossible- to draw conclusions from large volumes of data without the aid of mathematical analysis. In this particular case, our method allows for the automatic identification of subgroups of individuals, as well as the extraction of their common features, which represent their particular drinking behavior. Linking such characteristics through different phases, an evolution in the drinking behavior can be inferred. From this, one can analyze the small subset of features characterizing each pattern and draw meaningful conclusions.

Accordingly, we applied the developed a methodology to process time series data of the same group of individuals at different stages, identifying at each phase the underlying patterns of behavior and characterizing them in terms of a probabilistic model. A contingency analysis establishes the link between patterns of different phases, stating a pathway of the development of a final behavior.

Upon analyzing the observed patterns of behavior at each stage, (as well as the presence of dependency related symptoms, like ADE in spite of the aversive taste of the alcoholic solutions) we draw the following conclusions.

Intake measure and ADE classification

The first step on the analysis of data had to deal with a four-bottle paradigm (see Section 3.1). While in the literature the two-bottles paradigm is widely used [SS67, SL89], a four-bottle paradigm leads to a higher self-administration of alcohol [SH99]. However, the analysis of the alcohol consumption with the aid of the standard measures (net EtOH intake and solution preference) has some difficulties (see Section 3.2). After a comparison between several ethanol intake measures, a water penalized net $EtOH$ intake is defined (see Section 3.2.3). This reflects several features that neither the net $EtOH$ intake nor the $EtOH$ preference present. While a reduction in the net $EtOH$ intake was observed, the water penalized net $EtOH$ intake remained constant throughout phases.

The new measure allows to explain how rats at advanced phases decrease the overall $EtOH$ intake when quinine is added (since the taste is very aversive), however, to compensate they drink less water, which affects the metabolism of the alcohol, probably keeping the blood alcohol concentration at the desired level in spite of the decreased intake.

A procedure for classifying animals into presenting ADE or not, based on the increased intake between baseline and after-deprivation phases, regardless of the presence of quinine or not is developed (see Section 3.3.2). The model of Sinclair et al. [SS67, SSJ73] of ADE was fitted to the ADE classified animals, yielding excellent agreement.

During the first after-deprivation phase, rats whose alcoholic solutions were altered with quinine were mostly classified as not presenting ADE. However, in later phases, most of them did present ADE regardless of the aversive taste on their solutions. These are not new results [SH99], and show how Wistar rats develop loss of control and inflexible drinking after several abstinence period.

Patterns of behavior

A characterization of drinking behaviors of Wistar rats throughout different baseline and after-deprivation phases of their drinking lifetimes was obtained. They reflect an evolution of the drinking behavior throughout baseline and deprivation phases. Each of the found patterns is characterized by the probability of a drinking event at each time point of the day, from each of the presented bottles (in this thesis, H_2O and 5%, 10% and 20% concentrated ethanol solutions). These results can be found in Chapter 6.

Section 6.2 contains the results obtained on baseline drinking patterns. During the first measured baseline phase, rats drink mostly in an explorative way: the more concentrated the alcoholic solution, the less they drink from it. Additionally they are characterized by a very high frequency of H_2O intake. In further phases, they develop a preference for a solution and drink with lower frequency, thus coming to a mature drinking phase. The net $EtOH$ intake reduces from baseline to baseline, however H_2O penalized net $EtOH$ intake remains almost constant throughout phases: They drink less alcohol but consequently also less water (or vice versa).

In Section 6.2, three patterns were found during after-deprivation phases. However their distribution over representation days varied from phase to phase. The more deprivation phases, the longer the intense after deprivation drinking pattern (ADP3) lasted, even in presence of quinine (from 5th after-deprivation on). The low after deprivation drinking pattern (ADP1) was always present after several days, as a sign of normalization after binge drinking days following abstinence. The second intense after deprivation drinking pattern (ADP2) tends to disappear throughout phases, which indicates a primary stage in withdrawal patterns. For detailed results on after deprivation patterns of behavior see Section 6.3.

Since on the latest phases most of the animals are classified as presenting ADE, no analysis of the risk of presenting an alcoholic addiction can be performed. However, as soon as new data sets are available, where the ADE outcome is not a fact, as it is in Wistar rats, important relationships between patterns of drinking and risks of developing an addiction could be uncovered following the proposed framework.

7.2   Discussion

In Chapter 6 we applied the methodology developed in Chapters 4 and 5 to data recorded under the *long term self administration with repeated deprivation phases* protocol for Wistar rats, which was described in Chapter 3.

On the methodological side, the results for multivariate data developed by [WD95] are generalized for the case of time series. These result can be seen as a further development on model based clustering time series, that could complete the methods reviewed in [FS11]. Here, a dynamic multinomial logit framework is described to model transition probabilities of inhomogeneous Markov chains. This can be seen as a special case of our proposed framework, by treating previous observations as covariates and assuming a multinomial distribution with parameter $n = 1$ for the transition matrix's rows. Many other specific applications can be generalized under the proposed framework, as long as they involve samples from a member of the family of exponential distribution. Our methodology gives a unified solution to a broad range of problems regarding the modelling of multivariate time series.

For the biological society, the proposed framework allows the modelling and evolution of patterns in many biological process. It allows not only the partition into meaningful groups, but also the dynamical evolution analysis throughout time of many processes, such as the before/after drug treatment change of targeted behaviors, and many other behavior-related analysis.

Our initial analysis on the overall ethanol intake showed that rats drink less net ethanol g per body kg in time. However, since water intake decreases as well, it can be seen that the fraction of ethanol intake per water (ml/dl) remains almost constant throughout phases. The H_2O *penalized net EtOH intake* proposed in Chapter 3 allowed the observation of this fact, and we therefore strongly recommend it in the context of a four-bottle paradigm as an intake measure.

Using the proposed intake measure, a classification of each individual in presenting ADE or not was developed. The net EtOH intake of the animals classified as presenting ADE were fitted to the model proposed in [SS67] obtaining very similar parameter values as the reported. We assume this as a confirmation of the validity of the proposed procedure.

The developed GLM-EM classification algorithm yielded a thoroughly dynamic characterization of ethanol drinking patterns during different stages of an individual's drinking lifetime. Patterns during different baseline phases reflect a clear evolution towards mature drinking behaviors, where the frequency of drinking events (water, as well as ethanol) is low, and there is a clear preference for a solution. The results concerning patterns of behavior after a deprivation phase showed a development of an acuteness of the intense drinking pattern (particularly ADP3, see Chapter 6), reflected in the fact that the more deprivation phases, the longer rats drink under this pattern during the first days of representation of alcohol.

The results presented in this thesis can not be directly compared to those obtained so far in humans. This is due to the fact that the analyzed data sets are very different (our data is automatically recorded every 5 minutes from in-cage rats, while most studies on humans are made on self reports on daily amounts of drinks), and so therefore are the obtained patterns. The proposed modelling approach reflects, however, the concept of stochastic drinking pattern described by Gruenewald et al. (see [GN94, Gru98, GRL$^+$02]). They propose a probabilistic distribution for a drinking event which is independent from the time of the day (data is recorded based on telephone-calls, where the individual reports how many "drinks" he drank during the day). We, on the other hand, have data finely describing drinking behavior which vary according to the time of the day. This allows us to establish the probabilities of drinking events based on the time of the day, and as it can be seen, they are dependent on the day/night cycle. Whether these results reflect the different human drinking patterns could be analyzed in the future and find possible correlations between them and either risk or even presence of an addiction.

7.3   Implications

From the technical side, a tool box for classification of time series into dynamical patterns, as well as their characterizations in term of probabilistic models, has been developed. It constitutes the unification of several results, namely the generalization of GLM for modelling time series [Li94] to modelling panels of time series, and the generalization of the results of [WD95] for fitting mixtures of GLM to multivariate time series. Its flexibility and generality allows the application of these tools to other data sets that follow a certain structure, namely that it consists of:

1. time series data obtained from several individuals describing a particular phenomenon,

2. observations at each time point distributed from one member of the family of exponential distributions,

   3. a final outcome/classification of each individual is provided that allows linking described
      behavior with the outcome.

Under this framework, the time series from point 1.) can be classified by means of our GLM-
EM algorithm into a small set of dynamic patterns describing the temporal behavior of each
subgroup of time series. The found patterns can be correlated to the outcome given by point 3.,
yielding a prediction tool for further realizations of the particular phenomenon. In this case, the
new time series is classified into one of the existing patterns, and a probability distribution of
the outcome can be given, conditioned on this pattern. Thus, we emphasize that this a general
methodology which can be easily applied to several fields of research.

From the biological side, the results presented in this thesis are consistent with many known
results on the evolution of drinking patterns in Wistar rats throughout their drinking lifetime,
including the inflexibility in drinking, given by the presence of ADE in spite of quinine. It fur-
thermore provides a detailed dynamic description of the patterns of behavior at each stage. The
identified patterns of behavior not only contain a circadian rhythm component, but also several
higher frequency ones, whose analysis could provide further insight in the process of intake and
metabolism of ethanol.

Furthermore, this thesis provides biologists with a methodology to classify individual animals
into presenting ADE or not, comparing the increase in ethanol intake (after a deprivation phase
with respect to baseline drinking levels) between controls and animals whose alcoholic solutions
contain a bitter substance, like quinine. Quinine-animals drinking at the same level as controls
can be identified as having developed an inflexible drinking behavior towards alcohol, thus a
dependency.

7.4   Limitations and recommendations

In this section we summarize the shortcomings of the methodology we developed and the data
used. We furthermore propose some ways to overcome them, giving place to future research
topics continuing the thesis.

From the biological point of view, the initial goal of predicting the development of inflexibility in
drinking is not applicable, due to the fact that most Wistar rats develop it after several phases
of deprivation. This is widely reported in the literature (e.g. [SH99]) and can also be observed in
Chapter 6. However, this limitation can be easily overcome by performing the same experiments
on rats known to develop this feature with a more random frequency (certain percentage of rats
does, and the rest does not).

From the technical point of view, a first limitation relates to the simplifications made in order
to obtain better computational performances. The assumption of a common behavior within a
subgroup ignores the individual variability within the subpopulation. This is, in our opinion,
not so critical, since our goal is to reduce the dimensionality of the data, so that conclusions can

be drawn. However, allowing for some individual variability in the model could be achieved by either adding *random effects* components [ZK91] or by allowing for over-dispersion within the GLM framework [Ait96].

The EM algorithm we developed has the limitation of converging to a local maximum (see [DLR77]). Because of this, EM has to be restarted several times to achieve a global maxima. We propose therefore to find good starting guesses, who can make EM converge to the global maximum. Several other methods have been developed for fitting mixtures of distributions which have a better performance, being however not so intuitive to understand and implement as EM. A future step in the developing of our toolbox is related to the implementation of other mixture fitting techniques (e.g. Gibb's sampler [CG92, ZK91]), comparing their performance and robustness, to choose the best one for our classification purposes.

7.5   Final comments

The results presented on this thesis provide experimental biologists with a tool for a better understanding of the dynamic characteristics of ethanol drinking patterns at different stages of an animal's drinking lifetime. It furthermore proposes a new ethanol intake measure, which quantifies relative net ethanol intakes with respect to water. This is, in our opinion, a much more meaningful measure as the traditional net intake or solution preference at least in the context of a four-bottle choice paradigm and for Wistar rats, which tend to decrease the net alcohol intake per body *kg*, decreasing however also the water intake and thus maintaining an almost constant level of EtOH *ml* per H_2O *dl*.

The methodology presented can be further extended to different applications relating to pattern discovery and description of time series datasets. One example of the this is related to micro-array time series analysis. In this case, the identification of group of genes with a common regulation throughout the day and/or with oscillating components, as well as the analysis of the found patterns in different groups of interest (e.g. sick/healthy) can be performed with the proposed methodology. This is a very challenging application, since the micro-array data sets are typically large, leading to computationally expensive analysis. However, the selection of a first smaller subset of genes with certain application-dependent properties (small variation throughout members of a group at each time point, etc.) with the help of some statistical tests, allows smaller size data sets, which can be accommodated, within the proposed framework.

Another example relates to fMRI image time series. Here three-dimensional images are analyzed in terms of single voxel[1] univariate time series. They are usually modelled in terms of linear regression models and the goal is to find regions that activate simultaneously in the presence of determined stimuli [BJS03]. Some fMRI time series clustering has been developed, based on K-means and hierarchical clustering [GTR$^+$99] and on finite mixtures of Pott models [XLW09]. Our framework could also be applied to this field to identify spatio-temporal patterns correlated

---

[1] a voxel is the generalization of a two-dimensional image pixel to a three dimensional image, i.e. is the intensity value at certain tridimensional point of the 3-D image

to the afore mentioned stimuli.

Appendix A

Simulation study

Many theoretical results on GLM and EM have been developed, which adapt straightforward to the proposed framework[NW72, DLR77, FT01]. This chapter presents a simulation study aiming to verify some theoretical properties. The experiment targets the following objectives:

- Analysis of the performance of the estimator for β of the multinomial-logit model proposed in Chapter 4 for multivariate time series, assuming at each time t they distribute $Multinomial(n = 20, \pi_t)$. The consistency and asymptotic normality of the estimator are explored.

- Analysis of the performance of the classifier for several amount of clusters (h = 1,...,7)

- Analysis of the performance of the amount of clusters selection through BIC

A.1 Consistency and asymptotic normality of the estimator

Consistency is a very important property which establishes that the more data employed in the estimation, the closer will be the estimation to the real value. We refer to the MSE-consistency also called strong consistency [BM11]:

Definition A.1.1. MSE-consistency of the estimator $\{\widetilde{\beta}^{(n)}\}$ of β A sequence of estimators $\{\widetilde{\beta}^{(n)}\}_{n\to\infty}$, is an MSE-consistent estimator if

$$MSE(\widetilde{\beta}^{(n)}) = \frac{1}{NExp} \sum_{e=1}^{NExp} (\tilde{\beta}_e^{(n)} - \beta)^2 \to 0 \text{ as } n \to \infty$$

$\square$

In order to assert the MSE-consistency of the estimator $\widetilde{\beta}$ developed in Chapter 4, multivariate $NExp = 100$ time series data sets $Y^{(n)} = \{Y_{it}^{(n)} \sim Mult(\pi_t, n)\}_{t=1\ldots288}^{i=1\ldots87}$ are simulated for $n = \{20, 40, 80\ 100,\ 150,\ 200\}$. Each $Y^{(n)}$ contains $N = 87$ time series each of length $T = 288$ and $R = 4$ channels (like in the real data set). $\tilde{\beta}^n$ is estimated for each $Y^{(n)}$, and finally the $MSE(\widetilde{\beta}^{(n)})$ is computed. Figure A.1 shows the MSE of each parameter $\widetilde{\beta}_p^{(n)}$. The larger the amount of data used to estimate β_p, the lower is the mean square error $MSE(\widetilde{\beta}_p^{(n)})$ for each parameter. Since the MSE tends to zero when the length of the simulated time series grows,
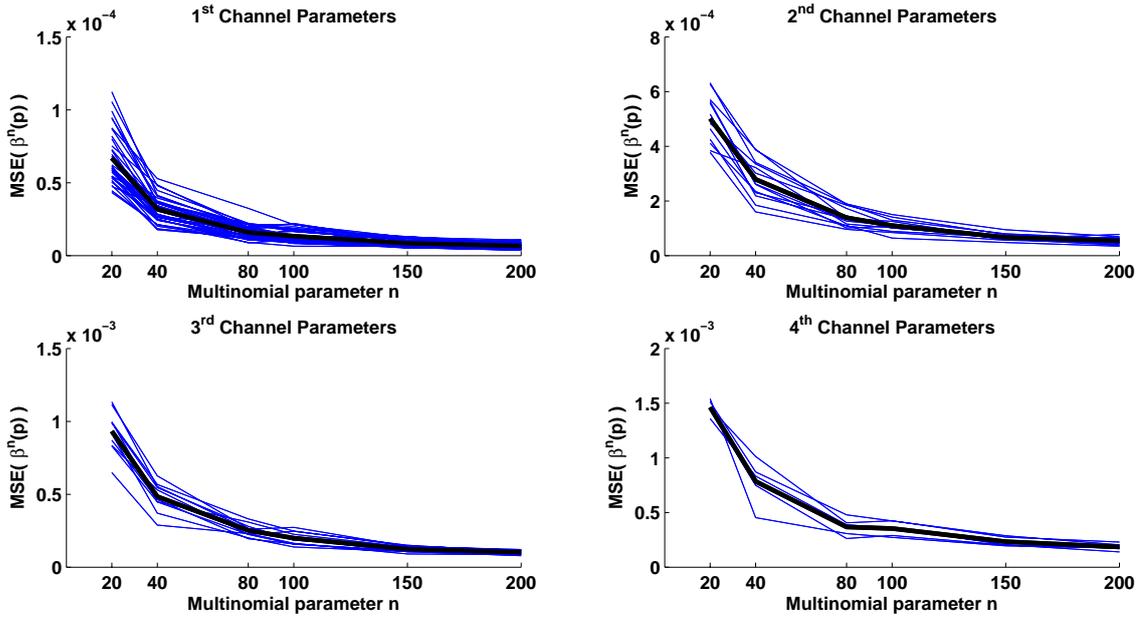
Figure A.1: $MSE(\widetilde{\beta}_p^{(n)})$: The larger the n, the smaller $MSE(\widetilde{\beta}_p^{(n)}) \Rightarrow \tilde{\beta}$ seems to be consistent.

the theoretical quantities should also have this property, thus it can be concluded that $\widetilde{\beta}^{(n)}$ is consistent.

A second important feature of an estimator is the

Definition A.1.2. Asymptotic normality

$$\sqrt{n * T}(\widetilde{\beta}^{(r,n)_p} - \beta_p^r) \xrightarrow{D} N(0, V), \text{ for certain } V < \infty,$$

for $r \in \{1, \ldots, R\}$ the output channels of the multivariate data, $n * T$ the amount of observations used to fit the data, and $p = \{1, \ldots P^r\}$ the parameter index for the r-th channel.

$\square$

Definition A.1.2 states that the distribution of the estimator takes asymptotically the shape of an normal distribution with fixed variance. Amongst the consequences of this property, it can be seen that the variance of the estimator tends to decrease in the rate of the amount of data used to fit the model. Figures A.2 shows in each row r the empirical distribution of the 100 values $\widetilde{\beta}^{(r,n)_p} - \beta_p^r$ (left) and $sqrt(nT)(\widetilde{\beta}^{(r,n)_p} - \beta_p^r)$. The solid red line on the right plots depicts an empirical distribution of $N(0, V)$ (obtained from a $N(0, V)$ sample of size 1000). On the left it can be seen that the variance of $(\widetilde{\beta}^{(r,n)_p} - \beta_p^r)$ decreases with increasing n, and on the right it is shown that it decays at a rate $\sqrt{n * T}$. This shows that our estimator, under the given assumptions, is asymptotic normally distributed. Though each row represents a single parameter $p = 3$ of each channel, the same holds for all the parameters (data not shown).
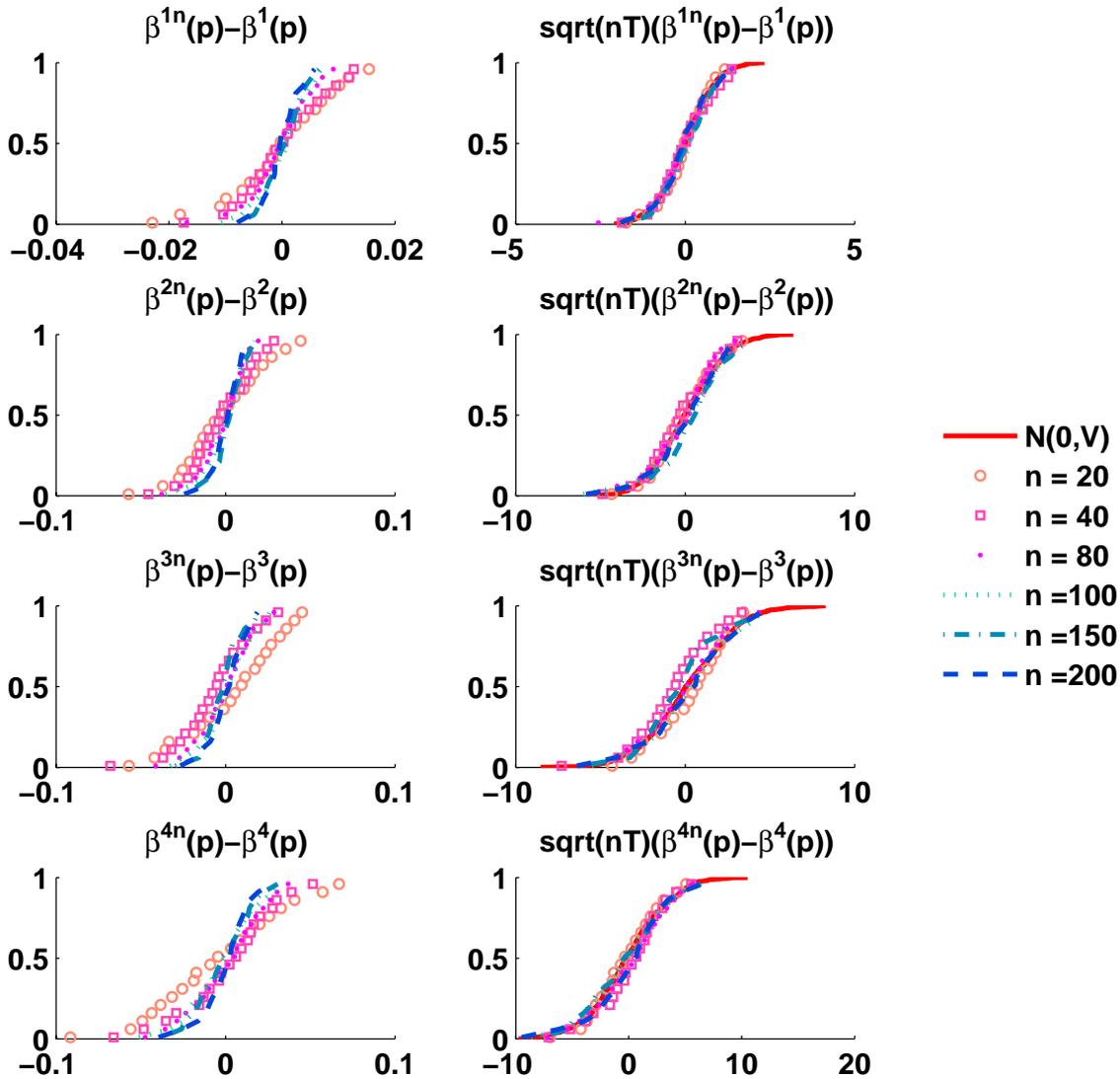
Figure A.2: Distributions of the values $\widetilde{\beta}^{(r,n)}(p) - \beta^r(p)$ and $\sqrt{n * T}(\widetilde{\beta}^{(r,n)}(p) - \beta^r(p))$ of 100 simulations for one parameter ($p = 3$) of each response channel (row-wise). To the left, it can be seen how the values $(\widetilde{\beta}^{(r,n)}(p) - \beta^r(p))$ tend to have a smaller variance with increasing time series length nT while in the right it is shown how the distribution variance decays with increasing n, i.e. $var(\widetilde{\beta}^{(r,n)}(p) - \beta^r(p)) \approx V/(nT)$, with a constant V for all n.

10^5*	Tested clusters (h')					Misc.
	1	2	3	4	5	$(h = h')$
Real clusters (h) 1	**1.266**	1.268	1.269	1.271	1.272	0
2	1.301	**1.258**	1.26	1.261	1.263	0
3	1.306	1.267	**1.243**	1.244	1.246	0
4	1.309	1.274	1.247	**1.229**	1.231	0
5	1.314	1.277	1.253	1.237	**1.226**	0.8%

Table A.1: Model selection on simulated data: BIC finds in all cases the right model.

A.2 Performance of the classifier and model selection through BIC

In order to analyze the performance of BIC in the model selection as well as of the classifier, an experiment is designed, where data is simulated following the structure of the original data. The general idea consists of fitting several mixtures, each of a different size, to the original data, and to simulate testing data sets from these fitted models. The programs are again ran for these data sets. Since the underlying models are known, several conclusions about the employed procedure can be drawn. In the following, the experiment is described in detail.

1. for $h = 1, 2, \ldots, H$ fit the multinomial-logit mixtures

$$F^h = \{\beta_1^h, \ldots, \beta_h^h, \pi_1^h, \ldots, \pi_h^h\}$$

   with h clusters to the original data Y.

2. for $h = 1 : H$ simulate $Y^h \sim F^h$, so that $size(Y^h) = size(Y)$.

3. for $h' = 1 : Y$, fit the mixtures

$$F^{h,h'} = \{\beta_1^{h,h'}, \ldots, \beta_{h,h'}^{h'}, \pi_1^{h,h'}, \ldots, \pi_h^{h,h'}\}$$

   and compute the log-likelihood $L(h, h') = log(L(Y^h | F^{h,h'}))$ and with this, the BIC.

If steps 2 and 3 are repeated several times, a mean $BIC(h, h')$ table can be computed and used to have a more reliable result on the model selection behavior through BIC.

Notice that the structure of each Y^h is known, so the misclassification rate can be computed for each experiment. The selected amount of classes through BIC can be compared with the real amount of classes used to simulate the data. This gives an idea of how good the clustering procedure works, given the stated distributional assumptions, and how good BIC discovers the underlying structure of the data set. Table A.1 shows these results.

A.3   Conclusions

The theoretical results of consistency and asymptotical normality of the estimator for GLM seem to hold when extended to mixtures of multinomial-logit time series. This was shown in Section A.1 with the help of simulated data.

The discovery of the underlying amount of clusters in mixtures of multinomial-logit simulated data was achieved through BIC, yielding the correct amount in all the experiments.

The classifier performs very good for the simulated data: it makes a very low rate of misclassification (less than 1%) throughout 100 experiments for each h' as shown in the rightmost column of Table A.1.

Further experiments should be undertaken, perturbing the assumptions with errors, which in the scope of this thesis has not been taken into account.

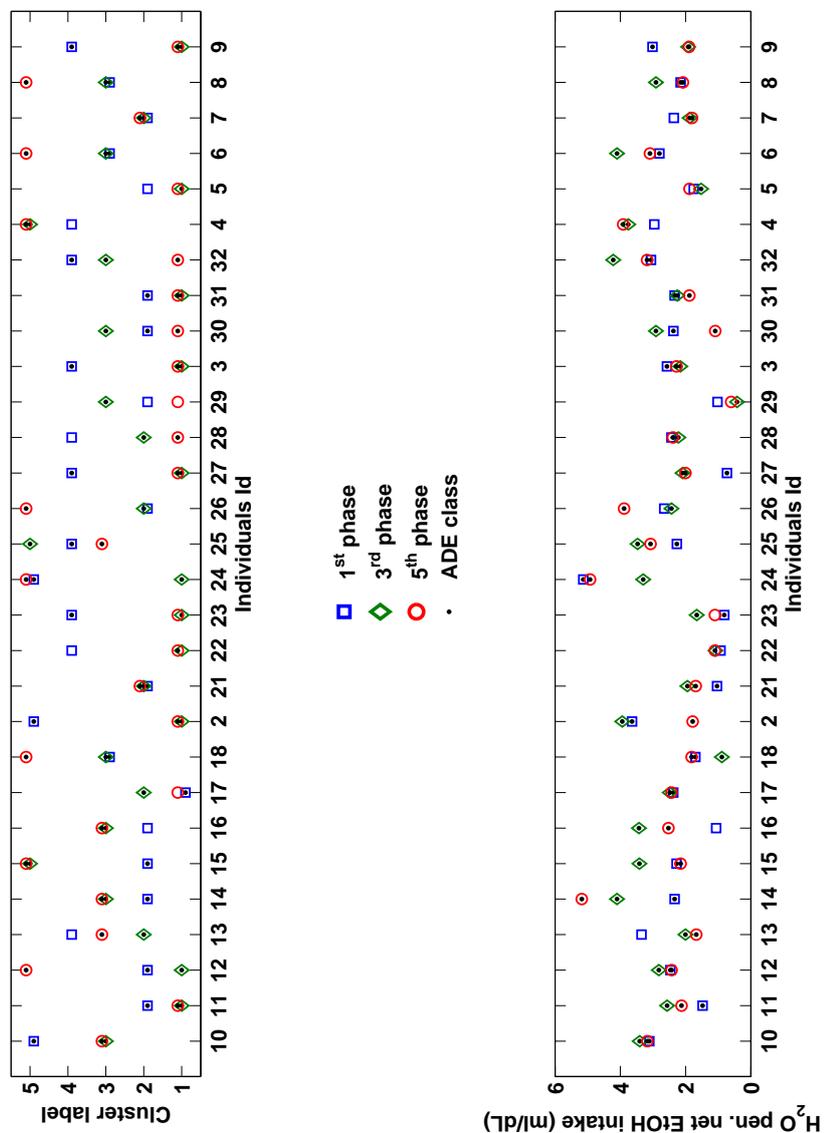# Appendix B

Some extra BP and ADP figures

Figure B.1: Classification of each individual on each baseline phase (top) and mean daily net
EtOH per baseline phase (bottom).
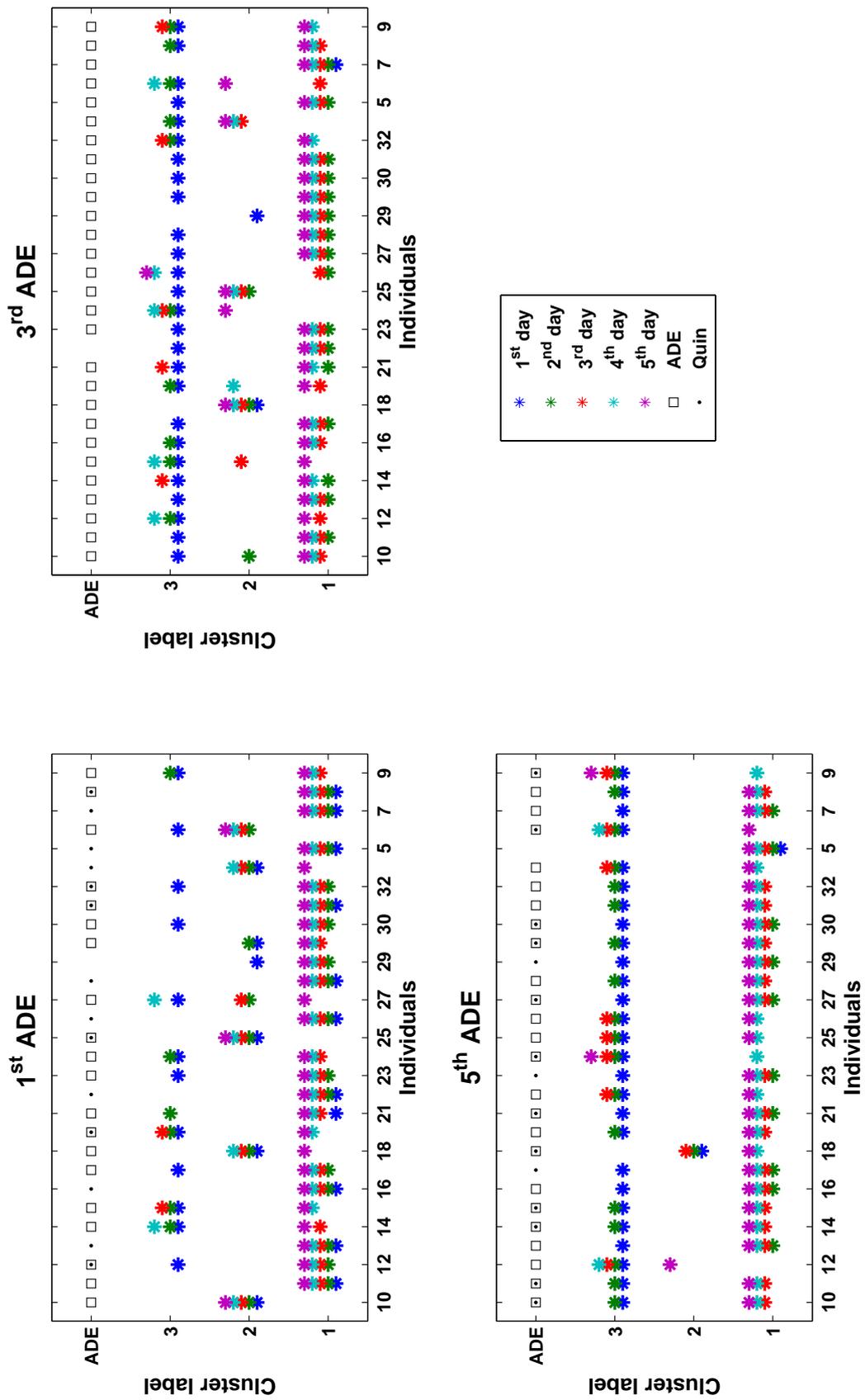
Figure B.2: AD phase - wise classification of rats throughout AD days

Figure B.3: Box plots of mean daily amounts of each solution per pattern. The clusters appear in the same order of their label, thus blue depicts BP1 while magenta represents the BP5.

Figure B.4: Box plots of mean drinking amount per drinking event of each solution per baseline pattern. The clusters appear in the same order of their label, thus blue depicts BP1 while magenta represents the BP5.

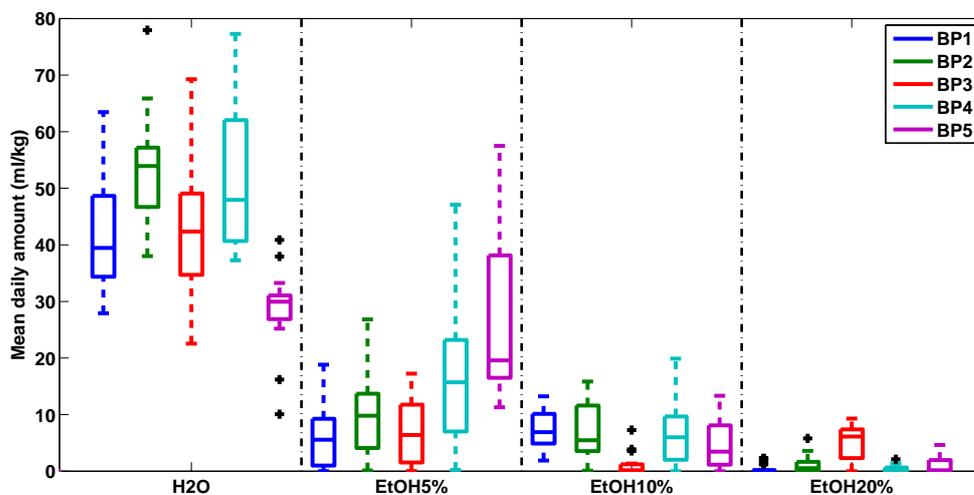Figure B.5: Box plots of mean daily amounts of each solution per after-deprivation pattern. The clusters appear in the same order of their label, thus blue depicts ADP1 while red represents the ADP3.
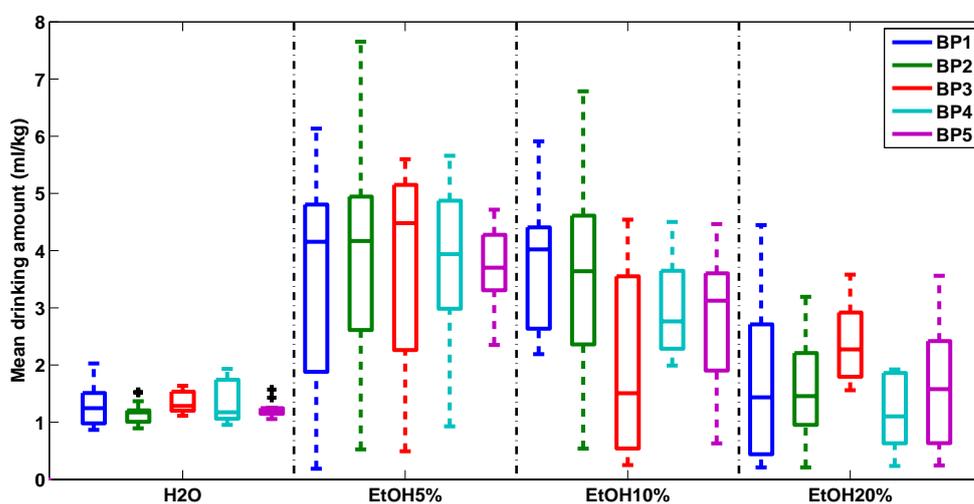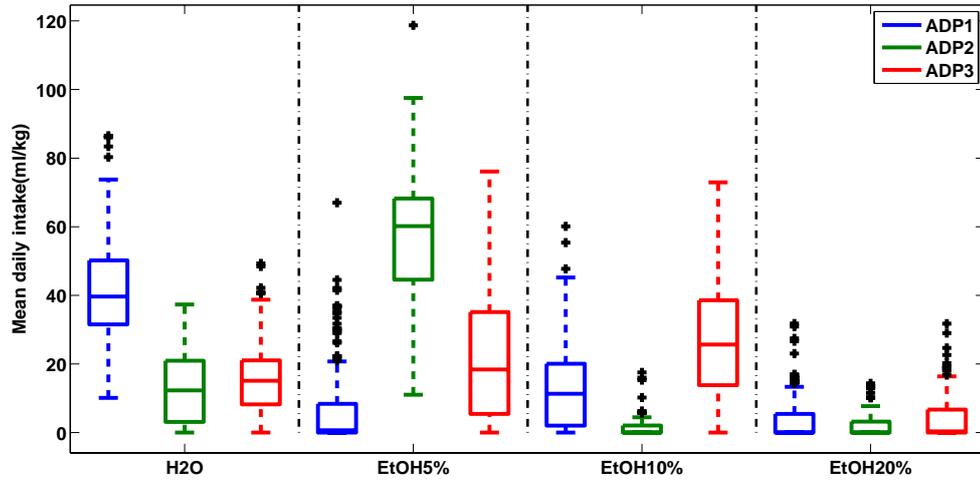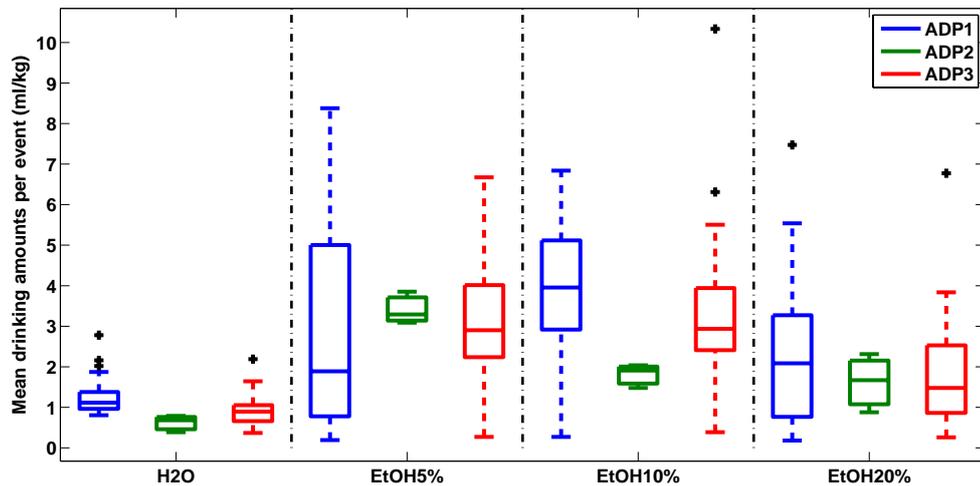

Figure B.6: Box plots of mean drinking amount per drinking event of each solution per after-deprivation pattern. The clusters appear in the same order of their label, thus blue depicts ADP1 while red represents the ADP3.

Bibliography

[ABH11]   Ch. Amann and J. Boysen-Hogrefe. A bayesian approach to model-based cluster-
            ing for binary panel probit models. *Computational Statistics and Data Analysis*,
            55(1):261 – 279, 2011.

[AHK42]   W. Ambrose, P.R. Halmos, and Sh. Kakutani. The decomposition of measures.
            II. *Duke Math. J.*, 9:43–47, 1942.

[Ait96]   M. Aitkin. A general maximum likelihood analysis of overdispersion in generalized
            linear models. *Statistics and Computing*, 6:251–262, 1996. 10.1007/BF00140869.

[And70]   E.B. Andersen. Sufficiency and exponential families for discrete sample spaces.
            Journal of the American Statistical Association, 65(331):pp. 1248–1255, 1970.

[Bev21]   W. H. Beveridge. Weather and harvest cycles. *The Economic Journal*, 31(124):pp.
            429–452, 1921.

[BG03]   S. Broadberry and B. Gupta. The early modern great divergence: wages, prices
            and economic development in europe and asia, 1500-1800. Department of Eco-
            nomics, University of Warwick, Coventry CV4 7AL, United Kingdom, September
            2003.

[BGJ73]   E.P. Box George and G.M. Jenkins. *Time series analysis: Forecasting and control*.
            Holden Day, San Francisco, CA, 1973.

[BJS03]   Christian F. Beckmann, Mark Jenkinson, and Stephen M. Smith. General multi-
            level linear modeling for group analysis in fmri. *NeuroImage*, 20(2):1052 – 1063,
            2003.

[Blo76]   P. Bloomfield. *Fourier Analysis of Time Series: An Introduction*. John Wiley and
            Sons, New York, NY, 1976.

[BM11]   W.R. Blischke and D.N.P. Murthy. *Reliability: modelling, Prediction, and Opti-
            mization*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2011.

[BMCS81]   T.G. Burish, S.A. Maisto, A.M. Cooper, and M.B. Sobell. Effects of voluntary
            short-term abstinence from alcohol on subsequent drinking patterns of college
            students. *Journal of Studies on Alcohol*, 42(11):1013–1020, 1981.

[Bol86]   T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal
            of Econometrics*, 31(3):307 – 327, 1986.

[BRL+06]   R.L. Bell, Z.A. Rodd, L. Lumeng, J.M. Murphy, and W.J. McBride. Review: The alcohol-preferring p rat and animal models of excessive alcohol drinking. *Addiction Biology*, 11(3-4):270–288, 2006.

[BRS+06]   R.L Bell, Z.A. Rodd, H.J.K. Sable, J.A. Schultz, C.C. Hsu, L. Lumeng, J.M. Murphy, and W.J. McBride. Daily patterns of ethanol drinking in peri-adolescent and adult alcohol-preferring (p) rats. *Pharmacology Biochemistry and Behavior*, 83(1):35 – 46, 2006.

[CAL+95]   G. Colombo, R. Agabio, C. Lobina, R. Reali, A. Zocchi, F. Fadda, and G.L. Gessa. Sardinian alcohol-preferring rats: a genetic animal model of anxiety. *Physiology and behavior, ISSN 0031-9384*, 57:1181–1185, 1995.

[CBM71]   W.H. Carter, J.V. Bowen, and R.H. Myers. Maximum-likelihood estimation from grouped poisson data. *Journal of the American Statistical Association*, 66(334):pp. 351–353, 1971.

[CdPBDF05] J.A. Chester, G. de Paula Barrenha, A. DeMaria, and A. Finegan. Different effects of stress on alcohol drinking behavior in male and female mice selectively bred for high alcohol reference. *Alcohol and Alcoholism*, 41(1):44–53, 2005.

[CG92]   G. Casella and E.I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):pp. 167–174, 1992.

[CHM+03]   I. Cadez, D. Heckerman, Ch. Meek, P. Smyth, and S. White. Model-based clustering and visualization of navigation patterns on a web site. *Data Mining and Knowledge Discovery*, 7:399–424, 2003. 10.1023/A:1024992613384.

[Cla05]   D.R. Clark. A Primer on the exponential family of distributions. 2005.

[CM73]   W.H. Carter and R.H. Myers. Maximum-likelihood estimation from linear combinations of discrete probability functions. *Journal of the American Statistical Association*, 68(341):pp. 203–206, 1973.

[CMB68]   T.J. Cicero, R.D. Myers, and W.C. Black. Increase in volitional ethanol consumption following interference with a learned avoidance response. *Physiology and Behavior*, 3(5):657 – 660, 1968.

[CPF+96]   J.C. Crabbe, T.J. Phillips, D.J. Feller, R. Hen, C.D. Wenger, C.N. Lessov, and G.L. Schafer. Elevated alcohol consumption in null mutant mice lacking 5-ht1b serotonin receptors. *Nature Genetics*, 14:98 – 101, 1996.

[CS73]   T.J. Cicero and B.R. Smithloff. Alcohol oral self-administration in rats: Attempts to elicit excessive intake and dependence. *Alcohol Intoxication and Withdrawal*, 1:213–224, 1973.

[CSPS71]   T.J. Cicero, S.R. Snider, V.J. Perez, and L.W. Swanson. Physical dependence on and tolerance to alcohol in the rat. *Physiology and Behavior*, 6(2):191 – 198, 1971.

[Del02]   F. Dellaert. The expectation maximization algorithm. 2002.

[DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):pp. 1–38, 1977.

[Eri72] K. Eriksson. Behavioral and physiological differences among rat strains specially selected for their alcohol consumption. *Annals of the New York Academy of Sciences*, 197(1):32–41, 1972.

[FK87] L. Fahrmeir and H. Kaufmann. Regression models for non-stationary categorical time series. *Journal of Time Series Analysis*, 8(2):147–160, 1987.

[FK98] K. Fokianos and B. Kedem. Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis*, 67(2):277 – 296, 1998.

[FS11] S. Frühwirth-Schnatter. Panel data analysis: a survey on model-based clustering of time series. *Advances in Data Analysis and Classification*, 5(4):251–280, 2011.

[FSK06] S. Frühwirth-Schnatter and S. Kaufmann. How do changes in monetary policy affect bank lending? an analysis of austrian bank data. *Journal of Applied Econometrics*, 21(3):275–305, 2006.

[FSK08] S. Frühwirth-Schnatter and S. Kaufmann. Model-based clustering of multiple time series. *Journal of Business and Economic Statistics*, 26(1):78–89, 2008.

[FT01] L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models (Springer Series in Statistics)*. Springer, 2nd edition, April 2001.

[GA76] D.B. Goldstein and V.W. Arnold. Drinking patterns as predictors of alcohol withdrawal reactions in dba/2j mice. *Journal of Pharmacology and Experimental Therapeutics*, 199(2):408–414, 1976.

[GN94] P.J. Gruenewald and N.T. Drinking in california: theoretical and empirical analyses of alcohol consumption patterns. *Addiction*, 89(6):707–723, 1994.

[Gol72a] D.B. Goldstein. An animal model for testing effects of drugs on alcohol withdrawal reactions. *Journal of Pharmacology and Experimental Therapeutics*, 183(1):14–22, 1972.

[Gol72b] D.B. Goldstein. Relationship of alcohol dose to intensity of withdrawal sign in mice. *Journal of Pharmacology and Experimental Therapeutics*, 180(2):203–215, 1972.

[GRL$^+$02] P.J. Gruenewald, M. Russell, J. Light, R. Lipton, J. Searles, F. Johnson, M. Trevisan, J. Freudenheim, P. Muti, A.M. Carosella, and T.H. Nochajski. One drink to a lifetime of drinking: Temporal structures of drinking patterns. *Alcoholism: Clinical and Experimental Research*, 26(6):916–925, 2002.

[Gru98] P.J. Gruenewald. Modelling the distribution and consequences of alcohol consumption. *Alcoholism: Clinical and Experimental Research*, 22(2):44S–51S, 1998.

[GTR+99]   Cyril Goutte, Peter Toft, Egill Rostrup, Finn Ã. Nielsen, and Lars Kai Hansen. On clustering fmri time series. *NeuroImage*, 9(3):298 – 310, 1999.

[GY20]   M. Greenwood and G.U. Yule. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83(2):pp. 255–279, 1920.

[Hal41]   P.R. Halmos. The decomposition of measures. *Duke Math. J.*, 8:386–392, 1941.

[Har58]   H.O. Hartley. Maximum-likelihood estimation from incomplete data. *Biometrics*, 14(2):pp. 174–194, 1958.

[Has69]   V. Hasselblad. Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, 64(328):pp. 1459–1471, 1969.

[HSW+02]   M.S.H. Henniger, R. Spanagel, A. Wigger, R. Landgraf, and S.M. Hölter. Alcohol self-administration in two rat lines selectively bred for extremes in anxiety-related behavior. *Neuropsychopharmacology*, 26:729–736, 2002.

[JGY91]   D. Jin-Guan and L. Yuan. The integer-valued autoregressive (inar(p)) model. *Journal of Time Series Analysis*, 12(2):129–142, 1991.

[JvdBGP98]   E.O. Johnson, M.B. van den Bree, A.E. Gupman, and R.W. Pickens. Extension of a typology of alcohol dependence based on relative genetic and environmental loading. *Alcoholism, clinic and Experimental Research*, 22(7):1421–1429, 1998.

[Kin18]   H.D. King. Studies on inbreeding. i. the effects in inbreeding on the growth and variability in the body weight of the albino rat. *Journal of Experimental Zoology*, 26(1):1–54, 1918.

[KMW99]   S. N. Katner, J.G. Magalong, and F. Weiss. Reinstatement of alcohol-seeking behavior by drug-associated discriminative stimuli after prolonged extinction in the rat. *Neuropsychopharmacology*, 20(5):471 – 479, 1999.

[LF73]   D. Lester and E.X. Freed. Criteria for an animal model of alcoholism. *Pharmacology Biochemistry and Behavior*, 1(1):103 – 107, 1973.

[Li94]   W. K. Li. Time series models based on generalized linear models: Some further results. *Biometrics*, 50(2):pp. 506–511, 1994.

[LQJ+98]   A. D. Le, B. Quan, W. Juzytch, P. J. Fletcher, N. Joharchi, and Y. Shaham. Reinstatement of alcohol-seeking by priming injections of alcohol and exposure to stress in rats. *Psychopharmacology*, 135:169–174, 1998. 10.1007/s002130050498.

[Mar60]   J. Mardones. Experimentally induced changes in the free selection of ethanol. *International Review of Neurobiology*, 3:41–76, 1960.

[Mel19]   E. Mellanby. Special report series no. 31. *London: Medical Research Committee*, 1919.

[Mel76]   N.K. Mello. Animal models for the study of alcohol addiction. *Psychoneuroendocrinology*, 1(4):347 – 357, 1976.

[ML98]   W.J. McBride and T. K. Li. Animal models of alcoholism: neurobiology of high alcohol-drinking behavior in rodents. *Critical reviews in neurobiology*, 12:339–369, 1998.

[MLE$^+$76] D. E. McMillan, J. D. Leander, F. W. Ellis, J. B. Lucot, and G. D. Frye. Characteristics of ethanol drinking patterns under schedule-induced polydipsia. *Psychopharmacology*, 49:49–55, 1976. 10.1007/BF00427470.

[MMR05]   J.-M. Marin, K. Mengersen, and C.P. Robert. Bayesian modelling and inference on mixtures of distributions. In D.K. Dey and C.R. Rao, editors, *Bayesian Thinking - modelling and Computation*, volume 25 of *Handbook of Statistics*, pages 459 – 507. Elsevier, 2005.

[MN89]   P. McCullagh and J.A. Nelder. *Generalized linear models (Second edition)*. London: Chapman & Hall, 1989.

[MP00]   G. Mclachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley-Interscience, 1 edition, October 2000.

[Mye62]   A.K. Myers. Alcohol choice in wistar and g-4 rats as a function of environmental temperature and alcohol concentration. *Journal of Comparative and Physiological Psychology*, 55(4):606–609, 1962.

[New86]   S. Newcomb. A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8(4):pp. 343–366, 1886.

[NW72]   J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):pp. 370–384, 1972.

[Osb92]   M.R. Osborne. Fisher's method of scoring. *Int. Stat. Rev*, 60:271–286, 1992.

[OVD09]   A. Owen, J. Videras, and L. Davis. Do all countries follow the same growth process? *Journal of Economic Growth*, 14:265–286, 2009. 10.1007/s10887-009-9046-x.

[Pea94]   K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of The Royal Society A: Mathematical, Physical and Engineering Sciences*, 185:71–110, 1894.

[PF03]   W. Penny and K. Friston. Mixtures of general linear models for functional neuroimaging. *Medical Imaging, IEEE Transactions on*, 22(4):504–514, 2003.

[Poh81]   L.A. Pohorecky. The interaction of alcohol and stress: A review. *Neuroscience and Biobehavioral Reviews*, 5(2):209 – 229, 1981.

[PP81]   S. Parker and F. Perry. A discrete arma model for nonlinear system identification. *Circuits and Systems, IEEE Transactions on*, 28(3):224 – 233, mar 1981.

[Pru93]   H. Pruscha. Categorical time series with a recursive scheme and with covariates. *Statistics*, 24(1):43–57, 1993.

[PS10]   Ch. Pamminger and Frühwirth-Schnattery S. Model-based clustering of categorical time series. *Bayesian Analysis*, 5(2):345–368, 2010.

[PW00]   D.B. Percival and A.T. Walden. *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, February 2000.

[Ric26]   C.P. Richter. A study of the effect of moderate doses of alcohol on the growth and behavior of the rat. *Journal of Experimental Zoology*, 44(1):397–418, 1926.

[RL74]   E. Rubin and Ch.S. Lieber. Fatty liver, alcoholic hepatitis and cirrhosis produced by alcohol in primates. *New England Journal of Medicine*, 290(3):128–135, 1974.

[Rob48]   H. Robbins. Mixture of distributions. *The Annals of Mathematical Statistics*, 19(3):pp. 360–369, 1948.

[RSC02]   M. Ramoni, P. Sebastiani, and P. Cohen. Bayesian clustering by dynamics. *Machine Learning*, 47:91–121, 2002. 10.1023/A:1013635829250.

[SH99]   R. Spanagel and S.M. Hölter. Long-term alcohol self-administration with repeated alcohol deprivation phases: an animal model of alcoholism? *Alcohol Alcohol*, 34(2):231–43, 1999.

[SHA$^+$96]   R. Spanagel, S.M. Hölter, K. Allingham, R. Landgraf, and W. Zieglgänsberger. Acamprosate and alcohol: I. effects on alcohol intake following alcohol deprivation in the rat. *European Journal of Pharmacology*, 305(13):39 – 44, 1996.

[Sin71]   J.D. Sinclair. The alcohol-deprivation effect in monkeys. *Psychonomic Science*, 25:21–22, 1971.

[SL89]   J.D. Sinclair and T.-K. Li. Long and short alcohol deprivation: Effects on aa and p alcohol-preferring rats. *Alcohol*, 6(6):505 – 509, 1989.

[Spa00]   R. Spanagel. Recent animal models of alcoholism. *Alcohol Research and Health*, 24(2):124–131, 2000.

[SS67]   J.D. Sinclair and R.J. Senter. Increased preference for ethanol in rats following alcohol deprivation. *Psychonomic Science*, 8(1):11 – 12, 1967.

[SS93]   R.M. Salimov and N.B. Salimova. The alcohol-deprivation effect in hybrid mice. *Drug and Alcohol Dependence*, 32(2):187 – 191, 1993.

[SSJ73]   J.D. Sinclair, S.Walker, and W. Jordan. Behavioral and physiological changes associated with various durations of alcohol deprivation in rats. *Quarterly Journal of Studies on Alcohol*, 34(3,Pt. A):744 – 757, Sep. 1973.

[Tau86]   T. Tauchen. Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177 – 181, 1986.

[WD95]   M. Wedel and W.S. DeSarbo. A mixture likelihood approach for generalized linear models. *Journal of Classification*, 12:21–55, 1995.

[WD03]   Greene W.H and Hensher D.A. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681 – 698, 2003.

[WL05]   T. Warren Liao. Clustering of time series data–a survey. *Pattern Recognition*, 38(11):1857 – 1874, 2005.

[XLW09]   J. Xia, F. Liang, and Y.M. Wang. On clustering fmri using potts and mixture regression models. *Conf Proc IEEE Eng Med Biol Soc*, 1:4795–8, 2009.

[ZK91]   S.L. Zeger and M.R. Karim. Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):pp. 79–86, 1991.

[ZZ04]   H.-T. Zhu and H. Zhang. Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):3–16, 2004.