UNIVERSITÉ DE LAUSANNE
FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

THE STATISTICAL ANALYSIS OF DYNAMICS AND COMPLEXITY IN PSYCHOLOGY: A CONFIGURAL APPROACH.

Thèse
présentée à la Faculté des sciences sociales et politiques de
l'Université de Lausanne pour obtenir le grade de
docteur en psychologie

par

Philippe Lemay

Lausanne
1999

UNIVERSITÉ DE LAUSANNE
FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

THE STATISTICAL ANALYSIS OF DYNAMICS AND COMPLEXITY IN PSYCHOLOGY: A CONFIGURAL APPROACH.

Thèse
présentée à la Faculté des sciences sociales et politiques de
l'Université de Lausanne pour obtenir le grade de
docteur en psychologie

par

Philippe Lemay

Lausanne
1999

À mes parents

To : EvilLive!@TheEnd.com

"Pourquoi ton temps est-il si réduit? C'est pour moi le véritable mystère. Envoie-moi la suite."

La suite, à son retour, sera déjà là, cela, il le sait avec certitude. Il n'est pas du tout anxieux d'arriver chez lui, de même que pendant la journée il n'y a eu aucun bonheur à cultiver cette attente, ni le désir de fréquenter la zone qui la contient. Le soir, il suffit d'ouvrir la lettre de même que le fichier joint. Il n'y a aucune peur, chez Timetolose, ni d'émotion excessive, rien que de la curiosité, ça oui, en même temps qu'une détermination imprévue. Il lit:

From : EvilLive!@TheEnd.com

"Est-il pour toi si difficile de comprendre, Timetolose? Si la raison pour laquelle il reste si peu de temps n'est pas encore claire pour toi, alors, ç'a été une erreur de m'adresser à toi. Je t'avais dit qu'il s'agissait d'une responsabilité, non d'un privilège. Je ne peux pas t'en expliquer les raisons, crois-moi, je n'y parviendrais pas. D'ailleurs, tout ce que tu as besoin est là, dans la fin de la nouvelle. Je suis sûre que tu vas la lire."

Daniele Del Giudice, L'oreille absolue, p. 97-98

Contents

1 Introduction **1**
1.1 The objectives of research . . . 2
1.2 General questions about dynamics . . . 2
1.3 Types of data . . . 5
1.4 Approaches to the study of change . . . 7
1.5 Delimiting the scope of this work . . . 9
1.6 Conclusion . . . 9

2 Configurations **11**
2.1 Selecting variables . . . 13
2.2 Reducing data scales . . . 14
2.3 Karnaugh maps: representations for configurations . . . 15
2.3.1 Analysis of a Karnaugh map . . . 18
2.3.2 More graphical Karnaugh maps . . . 20
2.3.3 Higher-dimensional Karnaugh maps . . . 22
2.4 Conclusion . . . 23

3 Preliminary analyses **25**
3.1 Given a sequence of observations... . . . 25
3.2 The transitional frequency matrix . . . 26
3.3 The transitional probability matrix . . . 28
3.4 Markov chains . . . 30
3.4.1 Types of Markov chains and Markov states . . . 32
3.4.2 Applicability of Markov chains . . . 32
3.4.3 First passage times . . . 34
3.5 Conclusion . . . 36

4 Graphical representations **37**
4.1 Exploratory data analysis . . . 39
4.2 Graphics theory and data displays . . . 39

4.2.1 Purposes of graphics 40
4.2.2 Principles of graphics theory 40
4.2.3 Limitations of graphics 41
4.2.4 Conclusion 41
4.3 Graphics for time series data 42
4.3.1 Time series plots for continuous variables 42
4.3.2 Phase spaces for continuous variables 44
4.4 Evolustrips: time series plots for categorical data 48
4.4.1 Univariate evolustrips 48
4.4.2 Multivariate evolustrips 50
4.5 State transition diagrams 52
4.5.1 Dynamical Karnaugh maps 54
4.5.2 Animated Karnaugh maps 61
4.6 Conclusion 63

5 Analyzing patterns of transitions **65**
5.1 Testing global transitional dependences 65
5.1.1 Auto-correlation coefficient of binary variables 65
5.1.2 Goodman and Kruskal's lambda 68
5.1.3 Chi-square and likelihood ratio statistics 70
5.2 Testing specific transitions 73
5.2.1 The test of proportions 73
5.2.2 Sackett's z-score for testing transitional probabilities 76
5.2.3 Chi-square based tests 78
5.3 Conclusion 80

6 Entropy and complexity **81**
6.1 Information theory: a primer 81
6.1.1 Entropy 82
6.1.2 Joint entropy 86
6.1.3 Mutual information 88
6.1.4 Conditional entropy 89
6.1.5 Relative entropy (Kullback-Leibler distance) 92
6.1.6 Conclusion 93
6.2 Information loss when reducing scales 94
6.3 Measures of complexity of dynamics 96
6.3.1 An index of complexity of transitions 97
6.3.2 Other complexity measures 100
6.4 Advantages of information theory 101
6.5 Limits of information theory 102

6.6 Conclusion . . . 102

7 Analyzing phases **103**
7.1 What is stationarity? . . . 103
7.2 Testing stationarity . . . 104
7.3 Moving-omnibus: a bifurcation-point detector . . . 111
7.4 Moving-entropy and moving-index of complexity . . . 113
7.5 Conclusion . . . 116

8 Higher order Markov chains **119**
8.1 Higher order computations . . . 119
8.1.1 Higher order transitional matrices . . . 119
8.1.2 The n-gram method . . . 121
8.2 Higher order dependence using auto-correlations . . . 122
8.3 Higher order dependence using chi-square statistics . . . 124
8.4 Index of complexity for past states influences . . . 125
8.5 Information theoretical methods for higher order . . . 127
8.6 Conclusion . . . 130

9 Modelling dynamics **131**
9.1 Types of models . . . 134
9.2 Estimating expected frequencies . . . 135
9.3 Loglinear models for three-way tables . . . 136
9.4 Goodness of fit and selection of models . . . 137
9.5 Example: modelling stationarity . . . 138
9.6 Example: modelling cross-dependence . . . 142
9.7 Discussion . . . 149
9.8 Conclusion . . . 149

10 Conclusion **151**
10.1 Summary . . . 151
10.2 Perspectives . . . 153

A ESM Empirical data **155**

References **157**

Summary

The recent trend in complexity theory brought many researchers to conceptualize phenomena as complex dynamical systems, self-organized and governed by strange attractors. It fundamentally renewed the traditional approaches for analyzing change that are longitudinal studies and ARIMA-like time series analysis. But psychologists who analyze data from such standpoint face enormous difficulties, partly due to a lack of an adequate mathematical background, constraining methodological requirements and limited availability of computer software.

We present in this thesis a fruitful alternative for exploring and analyzing time series data: the *dynamical configural approach* . Recognizing the inherently multivariate and coarse-grained nature of psychological data, we developed a coherent set of methods that graphically and statistically describe the dynamics of a system. The proposed methods integrate both the micro- (univariate) and macro- (multivariate) perspective.

Prototypical themes and questions related to the analysis of time series data are first reviewed: patterns of transitions, temporal (in)dependence, cycles, higher-order influences, phases, and complexity. Statistical methods dealing with each of these are detailed, exemplified with experimental data drawn from the author's previous researches. Novel graphical techniques such as Karnaugh maps and evolustrips were specially developed for the visual exploration of configurations. Statistical analyses are mainly performed through entropy-based statistics, drawn from information theory. Loglinear models, as represented in graphical models, allow to answer most of the time series questions. Fruitful perspectives on the analyses of the complexity and dynamics of psychological systems are finally discussed.

Keywords : time series analysis, complex dynamical systems, configurations, binary and categorical data, information theory, graphical modelling.

Foreword

The end of this twentieth century witnesses the emergence of new concepts and the beginning of an acknowledged synthesis in psychological research. Experiences of individuals result from active and circular interactions with their environment. Perceiving, thinking, feeling or doing strongly depend on multiple processes on biological, psychological and social levels (Morin, 1977; Dantzer, 1994; Clark & Isen, 1982; Nesselroade & Ford, 1984). This bio-psycho-social approach combines the systemic and cognitive-behavioral perspectives into a single framework.

Not only psychological phenomena are considered as bio-psycho-social, but are also viewed as *dynamical* . Psychologists long recognized that change is an intrinsic part of our daily life and that every psychological phenomenon should be considered as an ongoing *process* , not as a fixed *object* (Morin, 1977).

This point of view has been reinforced by the latest developments in mathematics, physics and other "hard" sciences, by what is mainly referred as the "complex dynamical systems" approach. Literature of the last 20 years has been completely invaded with concepts such as *chaos, self-organization, bifurcations, strange attractors* and the likes. It has penetrated the language and practice of psychology over the years (yet with a certain delay).

Psychologists tried to transfer these concepts into their domain. So far the use has been mostly metaphorical. James Gleick's book *The theory of chaos* (Gleick, 1987), an accessible account of the history of chaos theory, is mainly responsible for the large diffusion of the ideas and concepts of this emerging field. It certainly sparked the imagination of many psychologists and lead them to draw parallels between theories of chaos, complex dynamical systems, self-organization and their field of investigation. While analogies sometimes open up to well-defined research programs, they alone are not sufficient. A further step must be taken.

Researchers who attempted to empirically test the presence of chaos or self-organization in psychological processes were soon dumbfounded. Chaos-oriented analyses are based on quite complicated mathematical development, for which psychologists are usually not prepared. Very few clinicians master double derivatives and triple integrals underlying such analyses.

Another limiting factor to the empirical testing of chaos relates to methodological requirements. The huge number of data points and the range of variable scale (a true interval scale) is excessively demanding, rather difficult to meet in psychological researches. In ecological investigation how could psychologists ask subjects to answer ten thousand times questions having scales greater than a hundred points? Only psycho-physiological research can possibly meet the requirements. Psychological research, dealing with fewer data points and small-scale variables, seemed then to be left on its own.

But is it so? Is psychology condemned to study only structural phenomena, and not processes? Not at all. However a slight change of focus is necessary. Instead of obsessively searching for amount of chaos in data and the infamous butterfly effect in every psychological phenomenon, researchers are better off studying more relevant aspects of dynamics, such as *attractors, patterns of transitions and bifurcations points*.

How to describe dynamical patterns? How complex is the dynamic? How to spot bifurcation points? Such questions have traditionally been answered using linear models such as ARIMA and state-space models. But we were unsatisfied with this kind of approach. We wanted to go beyond linear equations and explore new ways of describing processes. But how? Are there other analytical methods for dealing with time series data? These questions drove the inquiry behind this thesis.

The methodological path

We were looking for a general methodology that would integrate various "constraints". It had to be a *bottom-up* approach, where we could start from empirical data instead from a preconceived top-down model; it had to simultaneously deal with *multiple variables* , individuals being viewed as multidimensional systems interacting with their environment; suitable *graphical techniques* were required in order to visualize the underlying dynamics; it had to be adequate with *coarse-grained* or small-scale data, or even better, with categorical data typically found in psychological researches; and finally we were interested in *nonlinear* models, because advances in complex dynamical systems theory showed profound limitations in traditional linear models.

We came across – and got indeed attracted by – the work of René Thomas (Thomas, 1979; Thomas & D'Ary, 1990), a Belgian chemist who developed *Kinetic Logic* , a methodology for modelling (chemical) processes using boolean equations. Based on theoretical knowledge about the phenomenon under scrutiny he builds a set of logical equations, computes the corresponding table of transitions and analyses what possible states the system may go through, and the consequences for the system's behavior. Attractors and bifurcation points are then identified. This stimulating approach is exclusively *top-down* , building first the set of equations and then assessing the behavioral consequences of the system. While the general framework is sound, we were dissatisfied with an exclusively top-down approach.

Building on Thomas and D'Ary's work, we gradually unfolded a methodology that integrated the constraints. We first developed a graphical representation for the synchronical and dynamical organization of empirical data: the *Karnaugh map* . The graphical exploration of data is of primary importance, whether on cross-sectional or dynamical data. Exploratory data analysis (EDA) greatly emphasizes it as a first step before any numerical analyses (Tukey, 1977; Wainer & Thissen, 1993). There are two types of graphics for continuous time series data: time series plots and phases spaces. Unfortunately they cannot be directly employed on categorical data. So we invented equivalents for this type of data: evolustrips and dynamical Karnaugh maps.

It is a promising technique that not only shows what configurations are the most and least frequent, thereby pointing to structural attractors of the system; this *multivariate state transition diagram* also represents in a snapshot the transitions of the system, revealing the real dynamical processes the system undergone through a period of time. The resulting picture powerfully suggests mechanisms of adaptation of individuals. Both techniques are described in full length in chapter 4.

This graphical representation was a fundamental first step towards the right direction, but it needed to be supplemented with more rigorous statistical analyses. Graphics, arrows and frequencies alone cannot fully grasp all intricacies of a system. Furthermore human beings - researchers included - are prone to numerous perceptual and cognitive biases (selection and generalization biases, self-fulfilling prophecies) such that the sole inspection of graphics should be avoided as much as possible.

Delving on information theory a simple to understand – and apply – algorithm computing the complexity of a system's transitions was re-actualized. It is known in mathematical literature as *the conditional entropy of order k* , but we nicknamed it the *index of complexity of transitions* . It measures the entropy or the variability of the patterns of transitions. The smaller the index the less variable the transitions of the system, and conversely. This computation soon became indispensable, because it strikingly summarized in a single number the dynamics parallelly shown on the Karnaugh map.

This index of complexity, a quantifiable measure of the system dynamics, lacked one important feature. It was only descriptive, and not inferential. One could not tell if a system exhibiting an index of 0.875 was indeed a complex system or not, and if it was truly different from another exhibiting an index of 0.925.

Further investigating literature we then found the perfect complement to our visual and descriptive configural techniques: sequential analyses. Mainly represented by Gottman and Kumar Roy's book *Sequential Analyses* (Gottman & Kumar Roy, 1990) well-known statistical tests such as chi-square statistics are used. Much to our surprise some of the methods also employed entropy for computing diverse parameters of dynamics. The loop seemed then to be closed. We had graphical representations, descriptive measures of dynamics and statistical tests for valid inferences.

But a profound dissatisfaction remained. It soon appeared that there wasn't any statistical software available that had the desired functions. None had implemented entropy- or information-based functions (let alone indexes of complexity of transition) nor transition matrices, Markov chains, and Karnaugh maps. So we had to put our hands on the computer keyboard and program these functions ourselves.

At first there were attempts to program in Pascal language Karnaugh maps and the index of complexity. The programs served nicely to illustrate and analyze the dynamics and complexity of our bio-psycho-social systems. But it was cumbersome to add or modify features. Then came S-Plus.

Despite its command-oriented approach [1] S-Plus soon appeared to be a very efficient statistical software for programming custom-made functions. After a short period of learning all the previously programmed functions were introduced in S-Plus and a dozen were more added.

Not only did we program tests described in statistical literature, but we also developed various graphical and numerical procedures for describing and analyzing dynamics of categorical time series data.

Another set of techniques we developed is yet a transposition of methods used on continuous data. Often continuous data are smoothed so to yield a less "bumpy" or erratic evolution of data. This is performed using a moving-average; as it name implies, data are averaged over a number of observations (a time window of a given length) and repeated over and over by displacing the window by one observation in time. Of course there is no such thing as averages on categorical data. So we developed similar techniques: *moving-mode* , *moving-entropy* and *moving-complexity* . Instead of computing the average of observations for a given time window, one computes the mode, entropy or index of complexity of transition and repeat the process for the following time windows. Graphically represented, one instantly sees the moments when the dynamics was *ordered* and *crystallized* on a particular state or configuration, and when the dynamics is more disordered or complex.

Furthermore we developed an iterative procedure that allows to detect *bifurcation points* . Researchers may use the well-known chi-square test to assess if the dynamics before and after a given moment is different or similar. The sequence of observation is typically split in halves at mid-point. But we suggest to iteratively move this splitting point so to search for moments when the dynamics differ and

[1]Since 1996, software maker Mathsoft has provided with Splus 4 a menu-driven package.

when they are similar.

By then, even if research is a never-ending process and work is never finished, we had at our disposal a quite complete and coherent set of methods for dealing with psychological data. This thesis is thus intended to provide researchers in psychology the results of these underlying methodological reflections.

The outline of the thesis

Let's review the topics developed in this work. The objective is to describe graphical and statistical methods for the analysis of time series data. We mainly emphasize a *configural* approach, that is, a multivariate categorical approach, but these methods are suitable for univariate categorical time series data as well, configurations being a "special" case of categorical data.

The introduction surveys the three main approaches for the study of change and processes. They are: longitudinal studies, complex dynamical system theory and time series analysis. We briefly examine their objectives and how they are answered. The relative independence of the three approaches is reflected by the very different questions they ask. A synthesis attempts to gather the main directions they point to, paving the way to the precise questions we will answer later on.

In the second chapter is discussed what we consider a new and exciting way for dealing with multivariate categorical variables: configurations. We show how configurations are build, how to select variables, how to dichotomize them, and how a measure helps to quantify the amount of lost information when reducing scales.

Necessary preliminaries for the analysis of categorical time series data are presented in the third chapter. Researchers start with a given sequence of observations that is first transformed into important mathematical objects: the transitional frequency matrix and the transitional probability matrix. Markov chains analysis, a general methodology for dealing with such matrices, is then introduced.

Recognizing the importance of visually exploring the structure and dynamics of a system, we devote the fourth chapter to the description of suitable graphical representations for time series data. Graphics, together with the use of computers, have contributed to the rapid expansion and development of a dynamical perspective on psychological phenomena. For categorical variables exist univariate phase spaces called *state diagram diagrams* . In this chapter we present two innovative techniques we developed, a time series plot named *evolustrip* , and a *multivariate* state transition diagram named Karnaugh maps.

Chapter five concentrates on the description of the various analytical techniques for dealing with the dynamics and complexity of multivariate qualitative systems, answering the questions outlined in the first chapter. We present methods for describing and testing patterns of transitions. The auto-correlation, Goodman and Kruskal's lambda, chi-square and z-score based tests are examined.

The sixth chapter is a mainly "primer" on information theory, a major reference for the analysis of categorical data. The concepts and computations associated with this theory are presented: information, entropy, mutual information, conditional entropy and relative entropy. Advantages and limits of this theory are discussed. The application of information theory in categorical time series is exemplified in what we call the "index of complexity of transitions". Other measures of complexity are also discussed.

Analyzing time series data requires a fundamental property of data: stationarity, a stable structure in the dynamics. We examine in the seventh chapter what is this property and how it is assessed. Researchers may want to determine if a dynamics is characterized by different *phases* , periods where the system exhibits different patterns of transitions. We present a bottom-up procedure that iteratively seek point of *bifurcations* . Techniques derived from moving-average procedures are also presented: moving-entropy and moving index of complexity further allow to understand dynamics at different periods of time.

Most analyses are concerned with first-order Markov chain processes, that is, when the state of a system depends on the last state (but only on the last one). A system may depend on the k previous states. We describe in chapter eight techniques that determine the "memory" of the system, how past states influence the present one. They mainly rely on information theory.

The last chapter is a general methodology that answer some previously asked questions. Loglinear models allow researchers to model transitions of their systems, and answer such questions as stationarity of dynamics, compare transitions of two or more subjects and assess cross-dependences. They are presented using the graphical modelling approach.

The conclusion summarizes the main methods presented through out this work. A synthesis of their benefits and limits is proposed. Various perspectives on how they could be further developed are discussed. Not all possibilities were previously presented, so promising future steps could easily be undertaken. Fruitful perspectives are outlined.

All techniques, should they be graphical or numerical, are illustrated with empirical data gathered from an Experience Sampling Method. The experimental procedure is described in appendix A. Data mainly refer to a single subject. It provides a prototypical example of the power and versatility of these techniques, and thus serves as a large-scale case study.

The main purpose behind this work was to write a thesis (and statistical routines) that we would have liked to have 5 years ago, so we would not have been obliged to write this thesis. If both text and statistical software had then existed we could have concentrated on more psychological questions. The lack of integrated references on categorical time series analysis – moreover on a configural approach - and appropriate statistical package led us to "fill the gap". We hope that this purpose is achieved and that researchers will find the text and programs useful for their particular research.

A note on statistical software

Without appropriate references and tools, research, the relentless pursuit of knowledge and discoveries is doomed. Psychologists need statistical methods and techniques that answer the questions they ask. So we programmed many statistical analyses in S-Plus. This powerful software package allowed the necessary ease of use and flexibility to program the desired functions. The new time series analyses routines being integrated in a standard software, it allows to continuously extend its functionality, unlike a purely dedicated package such as SDIQ (Bakeman & Quera, 1995).

Readers may be interested in testing for themselves these techniques on their datasets, or test the validity of our claims. For that matter, the routines are made available. Through the author's web site, all statistical routines can be downloaded and imported in the S-Plus software. Point your web browser at the following address, and follow the given instructions:

http://tecfa.unige.ch/˜lemay/thesis/

Routines were originally programmed in the S-Plus statistical package; they nevertheless can be transposed to other packages such as SPSS and SAS by interested researchers (and seasoned programmers). If suggestions, critics or bugs are found, please report them to the author:

e-mail address: Philippe.Lemay@tecfa.unige.ch

Chapter 1

Introduction

Time. Evolution. Dynamics. Process. Change.

Many concepts expressing the fact that life is movement. Should it be the earth revolving around the sun in a year, the day and night going in and out following a circadian rhythm, breath alternating between inspiration and expiration, neurons responding to the impulse of other interconnected neurons. Life is composed of various dynamical intertwined processes.

Man has built various instruments in order to *measure* the flowing time: solar clocks, hourglasses, pendulums, watches. He also derived systems, such as calendars, that *record* the time that came and the time that is coming.

Not only man is interested in tracking time itself, but he is also interested in knowing if other phenomena are influenced by this intangible force.

In this challenging adventure, the inquiring mind will need a few elements that will maximize his chances of discovery: a general objective, often stated in the form of questions and hypotheses, a well-defined set of measures he may take and some analytical tools, nowadays in the form of computer programs.

These elements are deeply intertwined and co-determined through out the research process. The questions researchers want to answer depend on the objective pursued, which in turn depends on fundamental presuppositions about processes (epistemological beliefs), theoretical orientations, and the like. Furthermore the set of questions one may answer is limited by methodological constraints such as the type and quantity of data that can be gathered.

While we do not delve into the relation between objectives and questions, we detail here the interplay between questions and data. These two factors mainly determine the choice of statistical techniques that can be employed. And conversely, the available techniques determine the type of data researchers may measure and constrain the set of questions to can be answered.

1.1 The objectives of research

There are 5 general objectives that research - in general and more specifically about processes - may attempt to achieve. They are (Kendall & Ord, 1990):

1. description
2. explanation
3. forecasting
4. control
5. modelling

These objectives are not completely independent from each other, for the explanation of a phenomenon relies in part on its description, its forecast requires a detailed explanation, and so on. But researchers may concentrate on one or the other aspect. Most important, the objective pursued will affect the tools and techniques employed for the analyses.

The two most frequent objectives are description and explanation. Description is most often an exploratory phase undertaken using graphical representations and statistical measures that are not inferential, while explanation involves precise hypotheses to be confronted and employs inferential statistical tests.

Modelling is the latest, broadest objective (Bossel, 1994). It requires that the descriptive and explanatory phases brought sufficient information and knowledge about the system, so to build a model that synthetically gathers the various variables in a coherent and parsimonious way.

Control is an objective rarely set in psychological research (for it brings important ethical considerations), and forecasting is just a little more frequent. We will not address these two objectives in this work.

1.2 General questions about dynamics

Since we are interested here in phenomena that evolve trough time, we may wonder what questions could scientists ask about dynamics and processes? Beside precise theoretical interrogations leading to hypothesis testing, there are general, "content-free" questions about dynamics that transcend particular domains. We review the prototypical questions researchers ask about time series data.

We put forward an integrated set of prototypical questions about dynamics, regardless of the type of data. The questions are not strictly independent from each other, somehow overlapping a little. They will be detailed afterwards.

1. How can the dynamics of a system be graphically represented?
2. What are the underlying trends or patterns of transitions?
3. Does the system state at a given time depend on the state at previous time? With what strength?
4. Does the system state at a given moment depend on the k previous moments?
5. How ordered or complex is the dynamic?
6. Are there cycles in the time series?
7. Does the actual state of the system depend on other variables at previous time?
8. Is the dynamics stable across time, or are there particular phases?
9. Can forecasts be accurately made? To what precision?
10. Can a model of the dynamics be build?
11. Can dynamics of systems be compared? How can their similarity be assessed?

Graphical representations

Not a question per se, graphical representations of time series form an important part of the analyses, and should be performed whatever the questions to be answered. Exploratory data analysis (Tukey, 1977) emphasized the use of graphical displays as a pre-processing phase of analyses.

Trends & patterns of transitions

Trends are the general tendencies of a variable, as expressed by an increase or decrease in its levels. An information of paramount importance, this component of a continuous time series is often remove before analyses (by either smoothing or differentiating (Kendall & Ord, 1990)). A trend implies that data is not *stationary* , that is, that the structure of dynamics does not remain constant across time. Stationarity is a fundamental prerequisite for all time series analysis.

If there is no such things as *trends* in categorical time series (there are no quantitative levels in this type of data), researchers focus on *patterns of transitions* . This is really the most fundamental and general level of description. For example, one may want to know if an individual tries to maintain a low stress level in his daily life; the places and people an individual encounters after work; if an infant looks at his mother after seeing a new object; if sons follow the professional path of their father, and so on. These are patterns of transitions, for an event at a given moment is predetermined by events at previous moment(s).

Univariate state dependence

State dependence is the fundamental property of time series: state at a given moment depends on the previous state(s) of the system. If it were not depending on previous state, there would not be any dynamics. The whole purpose of time series analysis is to determine if there is such dependence, and how it is expressed.

Lag dependence

Lag dependence is the extension of the previous question, where one want to determine the higher order influences. Does the present state depend on the very last state? the two last states? the k last states? And what is the *strength* of the association between these past states?

Variability, order and complexity

Researchers should not only be interested in the *type* of patterns of transitions, but also on their variability. Since more than one type of transitions occurred, their *range* should be assessed. A related concept in dynamical system refers to the order and complexity of a system. There has been much debate about what how they are defined (Morin, 1977; Delahaye, 1994; Waldrop, 1992).

Cycles

Cycles are events or observations that repeat themselves after a certain period. They may occur rapidly, each second, day, month or year. The latter case is called the *seasonality* , in reference to the re-occurrence of seasons.

Multivariate state dependence

Multivariate state dependence, or temporal cross-dependence, concerns the dependence of the state of many variables at different times. The value taken at a given moment by a variables would not only depend on itself, but on the values of other variables as well, and at different lags.

Phases

Phases relate to the presence of a differentiated dynamics across time. It implies that the dynamics itself changed, a sort of meta-dynamics. Most analyses presuppose a non-changing dynamics, so there is a *methodological* reason for detecting the presence of phases. But there are also *conceptual* reasons, for researchers may indeed want to determine if there are phases; a patient may be subject to a clinical intervention at a given moment, and the purpose of the analysis is then to assess if the dynamics was modified by the intervention.

Comparisons of dynamics

Beside the description of the dynamics of an individual, it may be required to assess if two dynamics follow similar patterns of transitions or not. For example, in clinical settings, psychologists can determine if subjects which where given a certain treatment followed a different dynamics from a control group.

Modelling

Modelling corresponds to the most general analysis that one may performed. It implies building an abstract representation of how a system behaves. Its purpose is to assess the network of dependences among variables. It serves to answer most of the previous questions (phases, cross-dependence, comparison of dynamics, . . .)

Forecasting

Forecasting deals with the prediction of future states of the system. An uncommon task in psychology (yet common in economics), but nevertheless important. Given a researcher has produce a model of the system behavior, he may predict the future states that are likely to be encountered. For clinical purpose it may reveal necessary, after the introduction of a treatment. In what state (positive or negative) will patients be weeks or months afterwards?

1.3 Types of data

Once the questions to be answered have been specified, researchers proceed to transform the given observables into data. There exist various types of data, for which we distinguish between (Fienberg, 1980; Howell, 1997):

1. dichotomous or binary
2. nominal or non-ordered polytomous
3. ordinal or ordered polytomous
4. integer-valued
5. continuous

The first three types are said to be *categorical* or *qualitative* , because the values taken by variables can not be put on the real axis (in mathematical terms), which is the case for the last two types of data, said to be *numerical* or *quantitative* .

Dichotomous , or *binary* variables arise when there are only two categories: yes or no answer, failure or success, male or female, are examples of this type of data.

Nominal or *non-ordered polytomous* variables may take one of multiple values for which no order exists between them. Only names can be attributed to these categories. Examples are: marital status (single, married, divorced, separated, widow), citizenship (with names of country), and assignment to an experimental group (A, B or C) .

Ordinal or *ordered polytomous* variables exhibit an order between the various categories, but the distance between these categories can not be specified. One can not say that a category is 2 or 3 times higher (or bigger or better or whatever) than an other category. Some examples are social class (lower, middle, higher), diploma received (university, high school, . . .) and preference for a brand (none, little, moderate, high). Although categories of ordinal variables are sometimes assigned numbers and then treated as continuous variables researchers should be aware that this introduces bias in most analyses.

Continuous variables may take a high (infinite) number of values, which contrasts with variable types 1 to 4, said to be *discrete* , because they take any of a limited, exhaustive set of values. A further distinction is made between *interval scales* , for which the 0 value is arbitrary (for example, the Fahrenheit scale of temperature) and the *ratio scale* , for which the 0 is not arbitrary (variables such as age, height, reaction time are ratio-scale variables).

The distinction between the various data type is crucial for the subsequent analyses to be performed. Most statistical tests apply only for a particular type of data. Time series analysis for continuous data has constituted a well-developed body of statistical procedures on both conceptual and applied levels (by means of computer software) but not for categorical data. This discrepancy motivated us to write this thesis.

1.4 Approaches to the study of change

When designing the necessary empirical studies for analyzing the underlying processes, psychologists (and other scientists as well) may choose among three major approaches:

- *longitudinal studies*
- *time series analysis*
- *complex dynamical systems*

Longitudinal study is the typical orientation of psychological researches. "The use of longitudinal data and methods has recently become quite fashionable, and for many, longitudinal research is touted as a panacea for establishing temporal order, measuring change, and making stronger causal interpretations." (Menard, 1991, p. 3). Longitudinal data, such as collected by census, are as old as 1491 B.C.; in New France (Quebec) some have been periodically collected since 1665 and United States have census data from the first decade of its existence up to the present. Longitudinal studies serve two primary purposes: to describe patterns of change and establish the direction and magnitude of causal relationships (Menard, 1991). As opposed to *cross-sectional* studies, which assess variables only once, longitudinal studies, data are collected at least twice (or for two distinct periods of time), with the the same subjects remaining through the research. Analyses imply the comparison of observations between the different periods of time (Diggle, Liang, & Zeger, 1994).

They are mainly undertaken on a large sample of subjects and questioned only twice (pre- post-tests) or a very limited number of times. The main questions these researchers seek to answer relate to the change of *level* of some variables according to some *contextual factors* . Analyses are mostly performed using repeated measures analysis of variance.

The second perspective on the analysis of change is *time series analysis* . Contrary to longitudinal studies where a large number of subjects are assessed a limited number of times, time series analyses are studies of a small sample of subjects who answer very frequently (often more than a hundred times) a limited number of questions. The focus is on single subjects (a case study approach) and analyses are intended to discover individual patterns of transitions (rather than detect a mere change of level of some variable).

Two different methodologies exist from time series analyses, depending on the type of data. If data are of a continuous type, the preferred techniques are the regression analysis and the Box and Jenkins' ARIMA models (Box & Jenkins, 1976; Kendall & Ord, 1990; Ostrom, 1990). If data is categorical, the main techniques are sequential analysis (Gottman & Kumar Roy, 1990), or Markov chain analyses (Kemeny & Snell, 1976; Gottman & Notarius, 1978).

The ARIMA-like approach is well represented in statistical literature. It is the "de facto" reference of any time series analysis. Sufficiently developed statistical software (such as SPSS, SAS, S-Plus) include time series analysis modules that compute auto-correlations at different lags, test their significance and allow to build models using linear equations.

Unfortunately this abundance is far from being found when dealing with categorical time series data. References are scarce, and statistical software does not have explicit modules that deal with such data. Even if some analyses may be performed using chi-square statistics, psychologists cannot (easily) perform state transition diagrams, Markov chain analyses, test the significance of specific patterns of transitions, let alone measure their complexity.

The third perspective on the analysis of change, *complex dynamical system* , caused much a stir in the last decades. Originating from the work of physicists and mathematicians in the 1960's, it has since spread in other exact sciences, life sciences and human sciences (Gleick, 1987). Meteorologist Edward Lorenz showed that very small differences in initial conditions of systems may produce quite different evolutions. His discoveries were coined "chaos theory": a small effect may later produce large consequences. It is illustrated by the now famous "butterfly effect": a butterfly flapping his wings in Beijing may produce small perturbations that gradually amplify and cause hurricanes in New York weeks later.

Psychologists did not resist the temptation to transfer the evocative concepts of complex dynamical systems theory into their field of inquiry. The widespread use of the chaos theory concepts in psychology is mainly in a *metaphorical* manner. An der Heiden considers health and disease as a result of non-equilibrium dynamic (Heiden, 1992). Goldstein presents causality in the emergence of the new concepts (Goldstein, 1996);. Psychological aspects like communication (Aula, 1996), mental disorder (Bergeret, 1991), therapy (Parry, 1996), drug addictions (Skinner, 1989), creativity (Zausner, 1996) and consciousness (Combs, 1996) were all described as chaotic systems.

Other psychologists were convinced by the necessity to test those intuitions and bring numerical results. Work was directed either towards *empirical testing* of definite hypotheses or the *modelling and simulation of processes* . Tschacher (Schiepek & Tschacher, 1992) analyzed psychoticity in clinical settings. Dauwalder, Pomini and Lemay (Pomini, Lemay, Dauwalder, & Bersier, 1996; Lemay, Dauwalder, Pomini, & Bersier, 1996; Lemay, 1993) using a general bio-psycho-social model analyzed the dynamics and complexity of tolerance behavior. Pomini described the complex dynamics of self-control and rehabilitation (Pomini, 1997). Military leadership and the clinical consulting processes were analyzed (Bersier, 1997). Tschacher (Tschacher, Scheier,

& Grawe, 1998) assessed the change of order in psychotherapies. Other examples of empirical analyses of dynamical systems include keyboard pressing (Ward, 1996), odor perception (Freeman, 1991), limb movement (Kelso, 1990) and fractals in cardiac rhythms and physiology (Goldberger, Rigney, & West, 1990).

Trying to determine if a dynamics is chaotic requires a huge number of error-free observations of continuous type. Only psycho-physiological researches can meet those requirements. Indeed the sole examples of strange attractors stem from this area of research. How could psychologists ask subjects to answer 10000 times a questionnaire, with questions having a continuous scale and without the slightest error? Any reasonable person would say the enterprise is almost doomed to fail.

1.5 Delimiting the scope of this work

Because psychological data is often coarse-grained, having a "pseudo-quantitative" small scales (Likert scales of 6 or 7 points are often considered as interval variables, which introduces a bias in analyses; they should be considered as ordinal), methods drawn for complex dynamical systems and ARIMA-like methods are subject to severe caution. Specific data requires specific statistical analyses, and it is bad practice to twist the nature of data into another type.

Consequently appropriate methods are necessary to deal with psychological data of the binary, nominal or ordinal types. The scarcity of references on the analysis of processes for these types of data led to search for specific solutions. Beside chi-square based statistics and Markov chains analysis, there is a tremendous gap that awaits to be filled. And since researchers also needs tools, in the form of computer software to perform the analysis, this thesis has the objective to provide both text and statistical routines and procedures to help them achieve their endeavor.

Therefore for each of the questions previously described we will examine their underlying concepts, statistical tests and computer procedures that are suitable for categorical data. Because psychological phenomena are believed to be of a multidimensional nature, a novel approach that simultaneously integrating many variables, the *dynamical configural approach* , will serve as the backbone of our work. It is presented and developed in the next section.

1.6 Conclusion

The three approaches for the analysis of change, longitudinal studies, time series analysis and complex dynamical systems, have been briefly presented in this chapter. The typical questions of the time series analysis framework were reviewed. Each of them will be detailed in separate chapters, exemplified with empirical data.

Chapter 2

Configurations

Summary. *We strongly advocate a multivariate approach in the analysis of categorical times series data. Question arises if there is such a possibility with this type of data. How can researchers go beyond uni- and bivariate relationships among variables and include three and more variables in analyses? We answer the question by relying on configurations. We present in this section how configurations are build and represented.*

The building block of our strategy is the configuration. We define configurations as a multivariate combination of many variables having *emergent properties* (Morin, 1977; Haken, 1983); that is, properties not found at the micro (univariate) level. Such configurations are usually composed of 2 to 8 categorical variables, so to keep representations and analyses manageable and understandable.

Although we developed a dynamical configural methodology on our own, investigation of literature shows that configurations is already an established concept. *Configural frequency analysis* (Krauth & Lienert, 1973; Eye, 1990; Eye, Spiel, & Wood, 1996) is a method for analyzing the individual cells in contingency tables. Contrary to the usual contingency table analyses, which describe *relationships between variables* , Configural Frequency Analysis focuses on the *particular cells* , thus on the characteristics of subjects.

By concentrating on cells of contingency tables, Configural Frequency Analysis searches for *types* and *antitypes* . Types are cells exhibiting frequencies that are statistically more frequent from what would be expected given a particular model (usually a model of complete independence between variables). Antitypes are those cells that are statistically less frequent than expected (given the same model). Both types and antitypes bring to the researcher's attention exceptional combinations of variables, because they appear more or less frequently than expected.

One of the main advantage of the configural approach is that it makes no distinction between *dependent* and *independent* variables. Experience Sampling Method (ESM) researches seldom make such distinction (Delespaul, 1995; Brandsttter, 1991). Bio-psycho-social variables are so deeply intertwined, each of them influencing - and being influenced by - the other ones that it does not make sense to specify which variables come first in the global causal path. Configurations put them all on the same level.

Method Configurations are built by *appending* or *concatenating* variables into a single entity. We mostly work with binary variables here for ease of understanding and manipulation, although the same procedure may be applied to m-ary variables.

Using binary variables, any N-variable configuration has 2^N possible or theoretical combinations. That is, 3 variables have $2^3 = 8$ possible combinations, 4 variables have $2^4 = 16$ possible combinations, and so on. Using 3-value variables, there would be 3^N possible combinations; for example, 3 3-value variables yields 27 configurations and 4 3-value variables yields 81 configurations. Their graphical representations get then more complicated and a higher number of observations is required to ensure valid statistical tests. The more general case, where variables may not have the same number of values is obtained by letting $m_1, m_2, \ldots, m_N$ the number of values for the N variables. The possible number of combinations is m_1 * m_2 * ... * m_N.

The *configural states* [1], e.g. all values or states resulting from the concatenation of N variables, may be represented by a symbol consisting of the values appended together. Suppose one has three binary variables, x_1, x_2 and x_3. Since each variable may take a binary value 0 or 1, there are 8 possible configural states (000, 001, 010, 011, 100, 101, 110, 111). The order of the variables in the configuration has no importance, as long as it is preserved throughout the analyses.

For a matter of conciseness one may employ the equivalent of the binary number in decimal format. For example, using 3 variables the binary combination 101 is equivalent to 5 in decimal, since $1*2^2 + 0*2^1 + 1*2^0 = 5$. Because of the arbitrariness in the order of variables, the decimal number is not an indicator of any quantity. Configuration #6 (110) is not better, greater, or more than configuration #3 (011), whereas both have the same number of variables set to 0 (1 variable) and 1 (2 variables).

Building configurations require one to three steps, depending on the type of available variables and the objective of the inquiry:

1. selecting variables to be included in configurations;

2. reducing the original data scales, preferably in binary scale

[1]The term configuration refers to both the global construct resulting from the concatenation of the different variables and to any of the individual states. This equivocation should be dispelled by the context where the term is employed.

3. assembling the individual variables into a single entity

These three steps are reviewed in the next sections. But first let's examine a concrete example.

Example Through out this work we mostly describe configurations made up of three variables from our empirical ESM study. The following variables are dichotomized around the center of the scale: familiarity of situations (BFAM), emotionality (BEMOT) and stress (BSTRE), the B before each name standing for binary. The resulting configuration is exhibiting in table 2.1.

BFAM	BEMOT	BSTRE	Config.	Decimal	Verbal description
0	0	0	000	0	Unfamiliar situation, low emotionality, low stress
0	0	1	001	1	Unfamiliar situation, low emotionality, high stress
0	1	0	010	2	Unfamiliar situation, high emotionality, low stress
0	1	1	011	3	Unfamiliar situation, high emotionality, high stress
1	0	0	100	4	Familiar situation, low emotionality, low stress
1	0	1	101	5	Familiar situation, low emotionality, high stress
1	1	0	110	6	Familiar situation, high emotionality, low stress
1	1	1	111	7	Familiar situation, high emotionality, high stress

Table 2.1: Building configurations for BFAM, BEMOT and BSTRE

There are eight configural states in this configuration. Each may be referred to using either the binary concatenation or its decimal number counterpart. The corresponding verbal description is included as to provided a more explicit knowledge of the underlying characteristics of the configuration.

2.1 Selecting variables

The choice of variables to be included in configurations depends on the investigator's intentions. On the one hand he may want to investigate some specific theoretical questions; the choice of variables is then dictated by such endeavor. On the other hand, on a more exploratory stance, he may consider one of the following selection schemes:

1. select variables according to a **principal component analysis** or **factor analysis** (if variables are continuous or ordinal) or cluster analysis (on categorical data) performed on "static variables", thus from a structural point of view. It allows to gather variables that are meaningful on a "synchronical" level, and then explore how they dynamically organize over time.

2. select variables combinations for which the **mutual information** is high (cf. section 6.1.3 for more details). An interesting procedure, but the mutual information should not be too high, for it means that variables have a high degree of redundancy.

3. select variables having high dynamical cross-dependences; later we will see statistics describing how two variables are dependent from a transitional point of view. These include Sackett and Gottman statistics, an index of cross-complexity and of co-variations.

Researchers may adopt two *stances* when considering the aggregation of variables. One is to maximize the similarity between variables, in order to observe how they co-evolve across time. The stance is to select minimally interrelated variables, that is, maximizing their differences so to have independent variables, that would in the end give a more complete picture of the state of the system. Performing a principal component analysis before follows the latter stance, because the result yields independent components.

2.2 Reducing data scales

Working with full scale data is sometimes cumbersome because of the large number of transitions included. It is far easier to work with a smaller number of categories. We may decide to reduce the number of categories to 2 or 3. It seems at first a drastic operation on data, leaving no subtle variations. If the choice of the dichotomization process seems at first arbitrary and a source of radical reduction in information, a measure allows to quantify the amount of loss of information involved by the scale reduction; it is presented in section 6.2.

Method How to determine points of dichotomization and trichotomization is far from a trivial task. For the reduction to a binary scale, one may consider many options: a dichotomization at the *center, mean* or *median* of the scale:

1. *center* : preferable when a natural opposition is present in the scale. In our studies we mostly had bipolar scales indicating presence or absence of social support, presence or absence of material support, etc. So this was the method we employed.

2. *mean* : probably the best choice if there is no obvious way to cut the scale at the center.

3. *median* : at first a interesting procedure, but it artificially increase the entropy of the recoded variables (cf. chapter 6.1); if there are 50% of the observations in each of the two values, then the entropy becomes maximal.

Discussion Dichotomizing full scale data at their center or mean are probably the most typical cases. The former is undertaken when the full scale is bipolar and has an even number of points; binary values are then interpreted as one side *vs* the other side of the scale. The mean of the scale is used when the scale is uneven or unipolar; the interpretation for values 0 and 1 respectively states "below" or "above" average, or low and high. Finally the median cutting point, so to have 50% of observations on each side of the scale, brings an interpretation of the type "below" or "above" median; because this option artificially increases the entropy of the system it is not a recommended option.

An important limitation for dichotomizing at the mean or the median comes from the absence of direct *interpretation* . How to interpret the 0s and 1s of such variables? Researcher would have to specify every time that 0 means below the mean (or the median) of variable X and 1 means above the mean (or the median). Reports would become very complicated to read.

Another limitation arising from these two dichotomization is the *lack of shared references* . Different variables would be "cut" at different points would not have the same meaning throughout.

Conclusion This suggested step is so far only a *recoding operation* .The manifold benefits will become apparent in next section (2.3). Researchers who wonder what are the consequences of data scale reduction may read section 6.2 to analyze how much information is lost in the process.

2.3 Karnaugh maps: representations for configurations

Manipulating a large number of configurations is sometimes tedious. Plotting 8, 16 or 32 states takes a lot of place and is difficult to have an integrated view on how the various configural states are articulated.

A well-adapted data representation is a necessary step towards a better understanding of the underlying dynamics of psychological phenomena. For this purpose we use a technique of computer science that usually helps to derive logical functions from truth tables: the **Karnaugh map** (Leussler & Ham, 1979; Lagasse, Courvoisier, & Richard, 1977).

Karnaugh maps originate from the field of logic and computer science. It is a technique used to determine logical equations from truth table; its purpose consists of minimization of prime implicants. It is vastly employed for designing digital circuit (Lagasse et al., 1977).

Instead of writing a straight truth table, one disposes a matrix-like structure of binary values taken by each variable. On the rows and columns are placed the values of the independent variables and inside each cell is put the value of the dependent variable.

Here are the structures of two- three- and four-variable Karnaugh maps (figures 2.2, 2.3 and 2.4 respectively; inside the cells are written the binary combinations of variables).

	V2=0	**V2=1**
V1=0	00	01
V1=1	10	11

Table 2.2: Structural Karnaugh map of two variables

	V2=0 V3=0	**V2=0 V3=1**	**V2=1 V3=1**	**V2=1 V3=0**
V1=0	000	001	011	010
V1=1	100	101	111	110

Table 2.3: Structural Karnaugh map of three variables

	V3=0 V4=0	**V3=0 V4=1**	**V3=1 V4=1**	**V3=1 V4=0**
V1=0 V2=0	0000	0001	0011	0010
V1=0 V2=1	0100	0101	0111	0110
V1=1 V2=1	1100	1101	1111	1110
V1=1 V2=0	1000	1001	1011	1010

Table 2.4: Structural Karnaugh map of four variables

Using two variables is straightforward: it is just an ordinary contingency table. When using three and more variables, one has to put combinations of variables on rows and columns. For three-variable combinations one variable is placed on rows (one for each of its two binary values) and the two other variables are placed in columns. But here comes a little trick: on the Karnaugh map values do not follow the usual order 00, 01, 10 and 11; instead they follow the *Gray code* order, for which each combination differ from its neighbor by the change of only one value. Therefore it consists of 00, 01, 11 and 10. The same procedure is taken for 4 variables, except that on rows there are two variables and their 4 possible combinations.

Inside the Karnaugh map the sequence of configurations does not follow the usual progressive order; it is arranged so that any two adjacent configurations differ from each other by the change of only one variable. For example, neighboring configurations 100 (#4) and 000 (#0) differ only by the third variable. Moreover the disposition of the map makes the sides also adjacent (top with bottom, left with right); although "far apart" states 000 (#0) and 010 (#2), differ only on the second variable.

Inside each cell of the Karnaugh map are put the value of the "dependent" variable, the one for which a function is sought. On figure 2.1 there is a truth table for which a simple compact function must be found. The Karnaugh map representation on the right helps to discover it (finding the equation is a matter of groups of cells having a value of 1 by adjacent pairs, 4-tuple in line or in square, or even 8-tuple in a contiguous block, which makes half the Karnaugh map; logical equations are then formed by a collection of "AND" "OR" connectors using values of group cells. Although we do not go into details here, the boolean function of figure 2.1 is $X = V3 \cdot \overline{V4} + V2 \cdot V3 + V1 \cdot V3$).

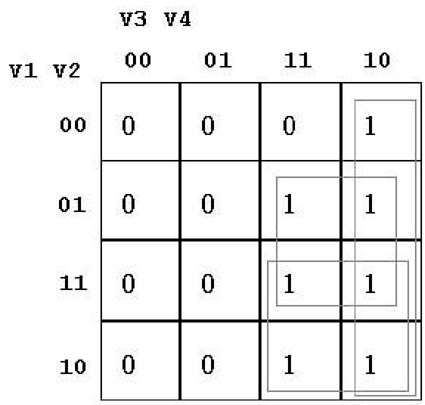

Figure 2.1: Example of a logical function described by a Karnaugh map

We did not intend to use Karnaugh map to find logical equations, although the *kinetic logic* approach (Thomas, 1979; Thomas & D'Ary, 1990) derives equations from it. Instead we had this simple yet powerful idea: to twist the primary use of Karnaugh map – a graphical technique for finding logical equations – into a graphical display of data. We first used it to plot frequencies of psychological configurations and then displayed its dynamics.

Let's illustrate the technique with configurations composed of familiarity of situations (BFAM), emotionality (BEMOT) and stress (BSTRE) for our subject. The usual linear arrangement of frequencies yields a table such as table 2.5. Integrated results would not be the main characteristics of this representation.

Config (bool)	Config (decimal)	Frequency
000	0	6
001	1	5
010	2	0
011	3	9
100	4	190
101	5	23
110	6	7
111	7	20
missing	99	1
Total		261

Table 2.5: Frequency of BFAM, BEMOT, BSTRE configuration

Let's now put the configuration frequencies in a Karnaugh map, as in table 2.6. The result is more compact and clearer to read. We will see how a *micro-macro* perspective emerges from such representation.

	BEMOT=0 BSTRE=0	**BEMOT=0 BSTRE=1**	**BEMOT=1 BSTRE=1**	**BEMOT=1 BSTRE=0**	**Total**
BFAM=0	6	5	9	0	20
BFAM=1	190	23	20	7	240
Total	196	28	29	7	260

Table 2.6: Structural Karnaugh map of BFAM, BEMOT, BSTRE

2.3.1 Analysis of a Karnaugh map

This mode of representation make its ideal for understanding both the univariate and multivariate structure of data. Configurations may be analyzed individually, *cell by cell* , to see what are the most and least frequent ones.

They may also be examined *variable by variable* : the disposition of the Karnaugh map makes it easy to do so. The first variable, BFAM, is displayed on rows 1 and 2; one sees that there are 20 cases of BFAM=0 (first line, 6+5+9+0) and 240 cases of BFAM=1 (second line, 190+23+20+7). It is rather clear that familiarity of situations predominates the experience scenery of this subject.

Now for the second variable, BEMOT. The value BEMOT=0 is obtained by the addition of the 4 cells of the 2 left columns: 6+5+190+23=224. The value BEMOT=1 is obtained by the addition of the 4 cells of the 2 right columns: 9+0+20+7= 36. More than anything this subject was in state of low emotionality.

Finally for the third variable, BSTRE. The value BSTRE=0 is obtained by the addition of the 4 cells of the 2 *border* columns (column 1 and 4): 6+190+0+7 = 203. The value BSTRE=1 is obtained by the addition of the 4 cells of the 2 *center* columns (column 2 and 3): 5+9+23+20=57. This subject thus reached more often states of low stress than those of high stress.

To understand the full power of the technique, here is a Karnaugh map of four variables. It is structured so that the first two variables are displayed on rows and the two other on columns, each combination in the map made up by their combined values. It should be again noted that the sequence of configurations does not follow the usual progressive order; it is arranged so that any two adjacent configurations differ from each other by the change of only one variable. For example, one observe that neighboring configurations 1101 (#13) and 1111 (#15) differ only by the third variable. Moreover the disposition of the map makes the sides also adjacent; states 0001 (#1) and 1001 (#9) are almost identical excepted for the first variable, having the value 0 for state #1 and 1 for state #9.

Illustrating the 4-variable Karnaugh map with a concrete example (see table 2.7), we get the following, using a fourth variable, BPRESS, describing perceived demands and pressures from the environment (BPRESS=0 is for low pressures and BPRESS=1 for high pressures).

Analyzing the map is proceeded as with the 3-variable Karnaugh map, except that there is a supplementary variable on the rows. Rows 1 and 2 represents cases where BFAM=0, while rows 3 and 4, cases where BFAM=1; rows 1 and 4 concerns PRESS=0 cases, and rows 2 and 3, PRESS=1. Rows totals make it clear that this subject experienced far more moments of low pressures ($n = 15 + 208 = 223$) than high pressures ($n = 5 + 32 = 37$).

We can observe from this Karnaugh map that there is a configural state that was very frequent, #1000 (n=171, which accounts for 66% of all encountered states): it consists of a familiar situation(BFAM=1), low pressure (BPRESS=0), low emotionality (BEMOT) and low stress (BSTRE=0). We may also observe that there are three configural states that were never experienced by this subject: #0100, #0010 and #0110.

	BEMOT=0 BSTRE=0	**BEMOT=0 BSTRE=1**	**BEMOT=1 BSTRE=1**	**BEMOT=1 BSTRE=0**	**Total**
BFAM=0 BPRESS=0	6	4	5	0	15
BFAM=0 BPRESS=1	0	1	4	0	5
BFAM=1 BPRESS=1	19	6	6	1	32
BFAM=1 BPRESS=0	171	17	14	6	208
Total	196	28	29	7	260

Table 2.7: Structural Karnaugh map of BFAM, BPRESS, BEMOT, STRE

It mostly implies that an unfamiliar situation and low stress are quite incompatible characteristics in a given situation.

2.3.2 More graphical Karnaugh maps

Putting frequencies or probabilities in a Karnaugh map table is a very fine way to understand the structure of some psychological data. There are more graphical extensions that make it even more informative, or at least, more communicative.

The most classical is the three-dimensional histogram, where on the x- and y-axes are placed the binary values of the independent variables and on the z-axis, the frequency of the configural states. Figure 2.2 is such graphics (same data as previously). The predominance of configuration #1000 is even more visible.

Other variations refers to a density plot. Here instead of letting the height of a bar to represent the frequency density plots use color or shades of grey. Darker colors mean more frequent configurations than lighter colors. The visual effect is as understandable as with traditional histograms. Figure 2.3 is such graphics.

We now describe a last variation that make a deeper parallel with our complex dynamical system conceptual framework. Remember the analogy of the marble in a bowl, or a rock in a valley. The steeper the mountain sides, the more chances rocks will take these ways (and the faster the rocks will go to the bottom). It also implies that it would require more energy in order to move a rock from a deep valley to another one. Attractors acts like valleys for rocks.

Histograms are transformed into valley-like graphics, by converting the direction of the bars (down, instead of up) and replacing bars by cones. The result (figure 2.4

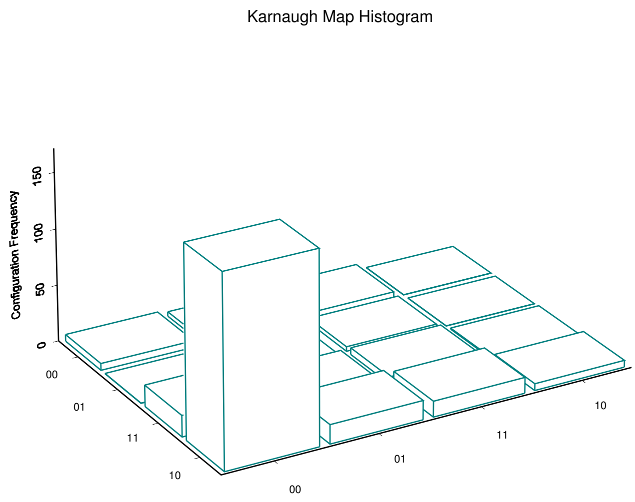

Figure 2.2: Karnaugh map histogram of 4-variable configuration frequencies

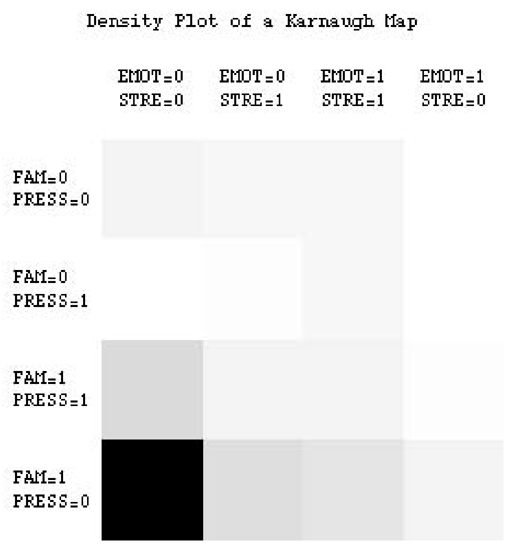

Figure 2.3: Karnaugh map density plot of 4-variable configuration frequencies

is very close to the idea one has about attractors. The more profound the cone (the more frequent the state) the more likely a subject is to be found in the state and the less likely he is to change its state.

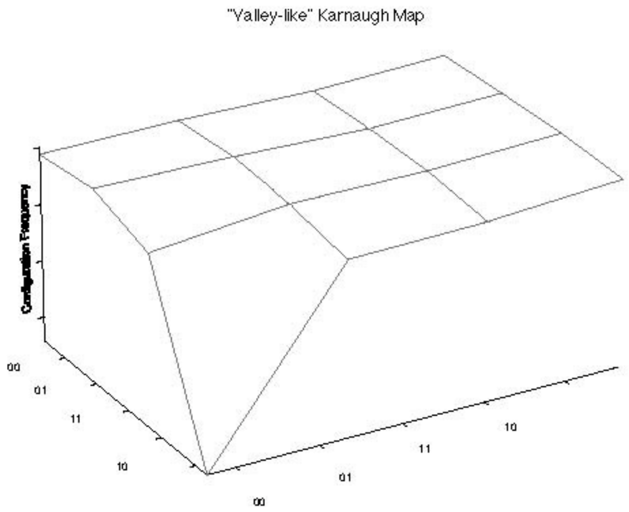

Figure 2.4: Karnaugh map cone of 4-variable configuration frequencies

2.3.3 Higher-dimensional Karnaugh maps

Three and four variables configurations Karnaugh maps will probably be the usual representations of data. One can also easily imagine five- and six-variable configurations. For configurations of 5 variables, one could just put side by side two 4-variable Karnaugh maps, where the left one is for v5=0 and the right one v5=1. The six-variable scheme would be to replicate the latter 5-variable Karnaugh map, on two "meta" rows: one meta-row would be the two previous 5-variable Karnaugh maps, for which v6=0; the second meta-row would represent cases of v6=1.

Anyone with a good vision for three-dimensional perspective would rather extends Karnaugh maps with stacked layers, instead of put them side by side. As an illustrative case, one could repeat the rows and column structure 00, 01, 11 and 10 on a vertical axis. This would provide the necessary space for a 6-variable Karnaugh map. No doubt this sounds exciting, but may not be that easy to understand. So for the moment we restrain ourselves to 3- and 4-variable configurations.

2.4 Conclusion

By presenting the configural approach, this section laid the foundations of subsequent analyses. Researchers know how to select variables, reduce data scales and build configurations. Karnaugh maps, at first a computer science technique for deriving logical equations, have been transformed into a graphical display of data frequencies. We describe in the upcoming chapters how Karnaugh maps are well-suited for the representation of configural dynamics and how to statistically analyze the dynamics of systems from a configural perspective.

Chapter 3

Preliminary analyses

Summary. *Given a sequence of observations, one builds a fundamental mathematical object, the matrix of transitions. It forms the basis of almost all subsequent analyses performed on categorical time series data. Matrices of transitions allow researchers to determine what are the general tendencies of variables. They are often – should always be – coupled with state transition diagrams. They may be applied on single variables as well as configurations.*

All methods dedicated to the analysis of categorical time series data - should they be sequential methods, Markov chain models and configural approaches - rely on a fundamental mathematical object: the transitional matrix.

3.1 Given a sequence of observations...

When considering a sequence of observations of a variable X, one sees transitions from one state to the other as a particular suite of symbols. How does one symbol follow an other? Is there a particular "logic", an organization behind the sequence? Transitional matrices are the first step towards an answer to this general objective.

Example Here is the sequence of observation, in raw form, of the stress level of a female subject (code name: FA7) who participated in our ESM study:

```
STRE = 4343255122542232123441444331322234222222525266333
4522322222222222222223322222245223225422332312324231412
315211233261332221111222211132421454656446222413422521
1423413533324463433313333233233655333333333443432313 1
111112212212223333323343333233343242232333334334433
```

A quick look at the sequence of observations let pop out an large number of states $STRE = 2$ and $STRE = 3$, and most important, a high number of sequence $2 \to 2$, and

$3 \rightarrow 3$. They seem to happen at given periods or *phases* of the sequence. This may be the sign of a *non-stationary* dynamic. This is explained in the following sections.

We shall not overemphasize the importance of a visual inspection of data, should it be in its raw form, like here, or using transformations. However this is obviously unsatisfactory here, the raw form does not lead to sufficiently valid inferences. Graphs of evolution (section 4) are better representations of the underlying dynamics. And the transitional frequency matrix (section 3.2) is the first and very fundamental operation performed on data that reveal information on the organization of the data sequence.

3.2 The transitional frequency matrix

The most basic question scientists ask about a dynamics is: when event a happens, what usually happens next? [1] Another event b or c? the same event a? This is the very essence of a process. A phenomenon that changes over time. What we want to determine is what event usually follows another event. But let's first see how a sequence of events may be described in a more compact fashion.

Method The sequence of observations is summarized by two fundamental mathematical objects, the transitional frequency matrix and the transitional probability matrix. The transitional frequency matrix is obtained by counting how many times did transition from state i to state j occur. The researcher counts the number of times that the sequence 1,1 occurred, sequence 1,2 and so on, until every possible sequence have been accounted for. He reports those frequencies in a square matrix $m \times m$, where m is the number of categories.

Example The transitional frequency matrix is a special form of contingency table. It has the general form given in table 3.1.

	$X(t+1)$				
	1	2	...	m	total
$X(t)$ = 1	n_{11}	n_{12}	...	n_{1m}	n_{1+}
2	n_{21}	n_{22}	...	n_{2m}	n_{2+}
...	...	...	...	...	...
m	n_{m1}	n_{m2}	...	n_{mm}	n_{m+}
total	n_{+1}	n_{+2}	...	n_{+m}	n_{++}

Table 3.1: General structure of a transitional frequency matrix

[1] *usually* is employed, since most of the systems we are dealing with are probabilistic, or *stochastic* . Well, they may be deterministic, but because the observations may be "stained" with measurement errors, or because not enough variables may be recorded at the same time, researchers may not grasp the whole complexity of the process. Therefore since every recorded event is not followed exactly by the same unique event, we are then obliged to consider the sequence of events as probabilistic.

For our subject example, where $m = 6$, the empirical transitional frequency matrix is displayed in table 3.2 (in bold are the highest frequencies for each row).

	$STRE(t+1)$					
	1	2	3	4	5	6
$STRE(t)$ = 1	**12**	9	7	4	1	
2	8	**45**	21	6	5	2
3	7	19	**42**	13	1	1
4	4	8	**10**	7	3	3
5	1	**5**	2	3	2	1
6	1	1	**2**	1	**2**	1

Table 3.2: Transition frequency matrix of stress

It should be noted that transitions are counted from state at time t to state at time $t + 1$. Why not counting from time $t - 1$ to time t? With empirical data, depending on whether one looks forwards or backwards, an observation is missing respectively at the end or the beginning of the sequence. But since missing values are typically not accounted for, the two types of transitional matrices are equivalent.

Analysis of transitional matrix There are plenty of relevant information about the dynamics of a system in the transitional frequency matrices. How are examined such matrices? What information is inferred from them?

The first step is to examine the *most frequent transitions* (abbreviated MFTs), that is, cells having the highest frequencies per row. Here most of the transitions occurred when stress was low, but not completely absent. The two most frequent transitions are from $STRE = 2$ to $STRE = 2$ (there were 45 such transitions), and from $STRE = 3$ to $STRE = 3$ (there were 42 transitions). The other MFTs are those $STRE = 2$ to $STRE = 3$ and $STRE = 3$ to $STRE = 2$. This strongly indicates a tendency for the subject to dynamically maintain a low stress level.

The second step is to examine the *least frequent transitions* (abbreviated LFTs), that is, cells having the lowest frequencies per row. The information is as relevant as the MFTs, because they represent *improbable* or *impossible* transitions, or even signs of *singular, decisive* transitions. Here there were few occurrences of very high stress ($STRE = 5$ and $STRE = 6$). Moreover there were not any transition from $STRE = 1$ to $STRE = 6$, meaning the subject never experienced a dramatic increase of her stress level; but she did experienced once a fantastic decrease, when her stress level dropped from 6 to 1. This strongly indicates a tendency for the subject to dynamically avoid a high level of stress.

Investigators are mostly interested in knowing what happened next when the subject was in a given state i; analyses of the transitional frequency matrix is done row by row, as will be the case through out this work. However an investigator may be interested in what *bring about* the occurrence of state j; she then analyzes the transitional frequency matrix column by column.

For our example one observes that the most frequent transition leading to a given state j is usually the same state: the highest transition to states 1, 2 and 3 at time $t+1$ are respectively state 1, 2 and 3 (with n=12, 45 and 42). However, for state 4 at time $t+1$ the highest transition was from state 3 (n=13), for state 5 is state 2 (n=5, indicating that moments of high stress were generally preceded by moments of low stress) and state 6 is state 4 (n=3). These last numbers may not be statistically significant, but nonetheless reveal a possible path of investigation.

Discussion The researcher may ask *why* those transitions exist. Were the most frequent transitions the result of an active process, a deliberate search of those preferred states? What amount of energy was spend to reach those states? And what about the least frequent transitions? They are usually considered as outliers in statistical literature, but may have been nonetheless triggered by significant events. Of course we can not tell if these LFTs are relevant or irrelevant; only a more detailed analysis of the context in which they occurred could separate the two opposite interpretations.

Conclusion Transitional frequency matrices are the first mathematical operation an investigator performs when analyzing the dynamics of categorical data. It reveals information about the underlying structure of the data sequence. One analyze the matrix by examining the most and least frequent transitions, either from a row by row or a column by column fashion.

3.3 The transitional probability matrix

Transitional frequency matrices are a nice method to explore the dynamics of a system. But when one wants to compare rows together, frequencies of transitions make the task difficult. A shortcut is to transform the transitional frequency matrix into a *transitional probability matrix* . The transitional probabilities are of the same nature than conditional probabilities: they represent the percent of transitions made from state i to j, *given* state i at time t.

Method A transitional probability matrix is obtained by dividing the frequency of transitions n_{ij} by the frequency of state i (n_{ij}/n_{i+}). It therefore represents a *conditional probability* , the probability to move to state j, *given* that the subject is in state i. The transitional probability should not be confused with the joint probability, which reflects the overall probability of observing a certain transition (it is computed by dividing n_{ij} by the total number of observations N).

Example The stress transitional probability matrix is given in table 3.3.

Analysis The analysis of a transitional probability matrix is undertaken row by row. As with a transitional frequency matrix, one may investigate the most and least frequent transitions, as accounted by the highest and lowest row percents (in bold typeface are the highest transitional probabilities).

For our example we see a strong tendency for this subject to return to states of low stress. Should she be in state $STRE = 1, 2, 3, 4, 5$, or 6, she has the highest probability to return to

	$STRE(t+1)$					
	1	2	3	4	5	6
$STRE(t) = 1$	**0.36**	0.27	0.21	0.12	0.03	0.00
2	0.09	**0.52**	0.24	0.07	0.06	0.02
3	0.08	0.23	**0.51**	0.16	0.01	0.01
4	0.11	0.23	**0.29**	0.20	0.09	0.09
5	0.07	**0.36**	0.14	0.21	0.14	0.07
6	0.12	0.12	**0.25**	0.12	**0.25**	0.12

Table 3.3: Transition probability matrix of stress

state of low stress (STRE=1, 2, 3). When STRE=6, she has an equal (but low) probability to reduce her stress to level 3 or 5. Examining the lowest transitional probabilities let us infer there are very few chances to transit to high stress levels.

Discussion There are limits to transitional probability matrix method. For less frequent states, transitional probabilities become "artificially" high. A transition $i \to j$ that is encountered twice is not the same if the state i was observed 3 or 50 times. But once again interpretation of these less frequent state is important: an "outlier" or a significant, critical event?

Another limit relates to the descriptive aspect of the transitional probability matrix. These are only "raw numbers". When confronted with transitional probabilities of 0.50 and 0.62, is it possible to conclude that the latter is really greater than the other? Is the difference significant? One must go beyond "face value" and statistically test hypotheses.

Conclusion Transitional frequency and probability matrices are the first mathematical operations an investigator should perform when analyzing the dynamics of categorical data. They reveal information about the underlying structure of the data sequence. One analyze the matrix by examining the most and least frequent transitions, either from a row by row or a column by column fashion.

This first step, as important as it is, remains an insufficient step. Investigators do not know if the overall dynamics is significant or not (cf chapter 5), or if one specific transition is significant (cf section 5.2).

Science being also a matter of communication, it is a necessary step to synthetically represent the transitional probability matrix information in a graphical form. State transition diagrams, evolustrips and Karnaugh maps answer this need; they are described in chapter 4.

3.4 Markov chains

Markov chains is one of the oldest and most widespread method for describing the dynamics of categorical systems [2] (Kemeny & Snell, 1976). They are a class of models describing the evolution of systems across time, referred to stochastic processes (those described by probabilistic functions).

According to Bersier (Bersier, 1997), who used this methodology to model change in orientation counseling, Markov chains have been used in psychological research for analyzing sequential events, decision making, problem-solving and social theorization.

The Markov chain framework allow investigators to answer different questions about the dynamics of systems:

[2] Markov chains also apply to continuous data, but it will not be treated here

1. If the system is in state s_i, what are the chances that it will be in some state s_j after some time t?
2. How much time does the system will spend in each state s, in average?
3. What is the mean and variance of the number of steps needed to go from s_i to s_j?
4. Are there any possibilities, that while going from a state s_i to s_j, it transits through another state s_k?

Investigators may also consider only a subset of states and examine how this subset behave. Markov chain theory provides a whole set of techniques to answer these questions, and is well suited for the analysis of categorical data.

A Markov chain may simply be thought as a process that moves from state to state. It starts in state s_i with probability p_i. It then moves to state another state j with probability p_{ij}. The transitional probability matrix is the basis on which all computations are performed.

Definitions and properties Let's first review the basic definitions and properties of these models.

The *first order Markov property* is that the state of the system at time t depends only on the outcomes of the last state:

$$Pr[X(t) = s_j|X(t-1) = s_{i_1}, X(t-2) = s_{i_2}, \ldots] = Pr[X(t) = s_j|X(t-1) = s_i] \tag{3.1}$$

A finite stochastic process is said to be an *independent process* if the present state of the system does not dependent on the past; that is, the probability of being in a state s_j at time t does not depend on the state s_i at time $t-1$,

$$Pr[X(t) = s_j|X(t-1) = s_i] = Pr[X(t) = s_j] \tag{3.2}$$

A finite Markov chain is a *finite Markov process* such that the transition probabilities $p_{ij}(t)$ does not depend on t. It means that at whatever time point t the chain is looked at, the transition probabilities p_{ij} are the same. This is the stationarity condition (it will be detailed and empirically tested in section 7.1).

The transition matrix for a Markov chain is the matrix P with transition probabilities p_{ij}. The *initial probability vector* for each state s need to be set; it is denoted as π_i and it often corresponds to the "static" probability distribution of the system, $Pr[X = s_i]$ (but this may not always be the case since one may force the system to start in a state s_i, thus $Pr[X = s_i] = 1$), or assign an initial probability vector.

The probability of being in state s_i at $t = 1$ is given by the product of the initial probability vector with the transitional probability matrix, $\pi(t = 1) = \pi(0)P(t = 1) = \pi(t = 0)P$, since P is the same for every t. The probability of being in state s_i at $t = 2$ is $\pi(t = 1)P^2$, and so on, yielding the induced relationship of equation 3.3. This means that to know the probability of being in state s_i at time t is obtained by multiplying the initial probability vector with the t-the power of the transitional probability matrix P (Kemeny & Snell, 1976).

$$\pi(t) = \pi(0)P^t \tag{3.3}$$

3.4.1 Types of Markov chains and Markov states

According to the properties of the ordering of states, different sets of states may emerge (Kemeny & Snell, 1976):

1. if it is possible to transit directly from any state s_i to s_j, the set of states is called *ergodic* ; once entered in an ergodic set, it is never left;
2. an *absorbing* state is one where it is not possible to leave it once encountered ($p_{ii} = 1$).
3. all other elements are called *transient* ; that is there is no possibility to stay in state s_i; once the system leaves a transient set, it never returns to it;

In every transitional probability matrix there is an ergodic set, consisting of at least one element, but not necessarily a transient set. If a chain as more than one ergodic set without a transient set, it means that there are no interaction between the two sets.

If the chain is ergodic without transient sets, it may be *regular* , where after a sufficient lapse of time the system could be in any state. It may also be *cyclic* , where the system encounters a series of different states and return to the original states after d steps.

If the chain has transient sets, the system will move to an ergodic set, either regular or cyclic; it cannot escape from an ergodic set once it enters one.

3.4.2 Applicability of Markov chains

The main presupposition for Markov chains analyses to be valid, is the process is of *first order* (the present state depends only on the past state) and that transitions are stationary (they do not depend on the period of observation). The former condition is tested in section 8 while the latter is examined in section 7.1

Example Let's find out what is the probability of our subject to be in a certain stress state s_i. We will use her transitional probability matrix shown in table 3.3. What is the initial probability vector? Let's pretend we do not know in what state she is now. The initial probability vector is then given by the "static" probability distribution (table 3.4).

State	**1**	**2**	**3**	**4**	**5**	**6**
Probability	0.13	0.33	0.32	0.13	0.05	0.03

Table 3.4: Initial probability vector (stress example)

What is then the probability to find the subject in a given stress state three time period ahead we observed her? We simply multiply the initial probability vector with the transitional probability matrix multiplied by itself three time (third power); If we compute the transitional probability matrix at a certain power k, it yields the probabilities for transiting from state i to state j to k times ahead. For a power of three, the transitional probability matrix is given in table 3.5:

So the probabilities to find the subject in a given state without knowing in what state she is now is (cf. table 3.6). The probabilities to find her in a given state are the same as with

	$STRE(t+3)$					
	1	2	3	4	5	6
$STRE(t) = 1$	0.13	0.34	0.32	0.13	0.05	0.03
2	0.13	0.34	0.32	0.13	0.05	0.03
3	0.13	0.33	0.33	0.13	0.05	0.03
4	0.13	0.33	0.32	0.13	0.05	0.03
5	0.13	0.34	0.32	0.13	0.06	0.03
6	0.13	0.33	0.32	0.13	0.06	0.03

Table 3.5: Transitional probability matrix at third power (stress variable)

the initial probability vector! It is not necessary so, but it happens when transitions converge rapidly to their long-term behavior.

State	**1**	**2**	**3**	**4**	**5**	**6**
Probability	0.13	0.33	0.32	0.13	0.05	0.03

Table 3.6: Probabilities to find the subject in a stress state, three times ahead

Strangely enough the transitional probabilities in table 3.5 are almost identical for each column. Why is it so? If we let the transitional probability matrix at a constantly greater power, the resulting matrix converges to its long-term behavior (the *limit matrix* , denoted by letter A). That is, if we let this person experience stress following a dynamics given by her transitional probabilities, the best guess about the state she is would be this limit matrix. In our example the limit matrix is (table 3.7).

	$STRE(t+\infty)$					
	1	2	3	4	5	6
$STRE(t) = 1$	0.13	0.33	0.32	0.13	0.05	0.03
2	0.13	0.33	0.32	0.13	0.05	0.03
3	0.13	0.33	0.32	0.13	0.05	0.03
4	0.13	0.33	0.32	0.13	0.05	0.03
5	0.13	0.33	0.32	0.13	0.05	0.03
6	0.13	0.33	0.32	0.13	0.05	0.03

Table 3.7: Limit matrix for the stress variable

The result implies that if we observe her for a large number of days we can expect her to feel 13% of very low stress, 33% of low stress, 32% of moderate stress, 13% of medium stress, 5% of high stress and 3% of very high stress.

Since each row of the limit matrix A converges to identical probabilities, it may be represented by a vector α. It may be algebraically computed by solving the equation: $\alpha P = \alpha$.

3.4.3 First passage times

Investigators may next examine how long it takes to transit from a state i to another state j. In Markovian terminology, this is called the *mean first passage time* .

The mean first passage matrix M is given by equation 3.4 (Kemeny & Snell, 1976):

$$M = (I - Z + EZ_{diag})D \tag{3.4}$$

where I is the identity matrix (a matrix containing ones in the diagonal and zeros elsewhere; see table 3.8), Z the fundamental matrix, E a matrix containing ones everywhere, Z_{diag} is the matrix containing in its diagonal the components of the fundamental matrix (and zeros everywhere) and finally, D contains in its diagonal $1/\alpha_i$ (1 divided by the components of the limit matrix).

I	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1

Table 3.8: Identity matrix for a 6x6 matrix

Z, the fundamental matrix, is computed by formula 3.5 and the numerical result of the stress example is shown in table 3.9.

$$Z = (I - (P - A))^{-1} \tag{3.5}$$

	$STRE(t+1)$					
	1	2	3	4	5	6
$STRE(t)$ = 1	1.33	-0.07	-0.16	-0.02	-0.04	-0.04
2	-0.05	1.26	-0.12	-0.09	0.01	-0.01
3	-0.06	-0.16	1.26	0.04	-0.06	-0.02
4	-0.01	-0.15	-0.05	1.09	0.06	0.07
5	-0.07	0.04	-0.25	0.09	1.13	0.06
6	0.00	-0.27	-0.13	0.03	0.25	1.12

Table 3.9: Fundamental matrix of stress variable

We can finally compute the mean first passage of this stress example (table 3.10). As expected the mean first passage is quite similar for each column. When the subject is in any of the six states, it takes about 10 transitions before she feels a very low level of stress (STRE=1); when she already feels such state, the average number of transitions is 7.89. It takes between 3 and 4 transitions before she experiences a low or moderate level of stress (STRE=2 or STRE=3), a much shorter time than for state STRE=1. For the other states, the average number of transitions increases, indicating that these states of higher levels of stress are more unlikely. The mean passage time is around 8 transitions for STRE=4, between 16.5 and 22 for STRE=5 and between 32 and 39 when STRE=6. So we conclude that on the average she is more likely to experience states of low and moderate levels of stress.

	$STRE(t+1)$					
	1	2	3	4	5	6
$STRE(t)$ = 1	7.89	3.96	4.38	8.46	21.82	38.34
2	10.85	2.99	4.25	8.98	21.01	37.28
3	10.92	4.23	3.08	8.01	22.26	37.65
4	10.55	4.20	4.05	7.66	20.09	34.56
5	11.02	3.63	4.67	7.63	18.73	34.81
6	10.47	4.56	4.29	8.10	16.49	32.89

Table 3.10: Mean first passage matrix for the stress variable

The *variance* of the mean passage time may be computed by a more complicated formula (Kemeny & Snell, 1976):

$$M_2 = M(2Z_{diag}D - I) + 2(ZM - E(ZM)_{diag}) - M^2 \tag{3.6}$$

We compute the standard deviation of the mean passage time (a much more easier to interpret measure of dispersion) by taking the square root of the variance; for our stress variable, it is displayed in table 3.11.

	$STRE(t+1)$					
	1	2	3	4	5	6
$STRE(t)$ = 1	9.74	3.56	3.77	7.94	20.82	36.56
2	10.32	3.24	3.74	7.97	20.79	36.56
3	10.32	3.64	3.37	7.88	20.82	36.56
4	10.30	3.62	3.71	7.83	20.72	36.43
5	10.32	3.49	3.78	7.84	20.55	36.44
6	10.29	3.62	3.78	7.86	20.10	36.24

Table 3.11: Standard deviation of mean first passage matrix for the stress variable

Like the mean first passage time its standard deviation depends very little on the choice of the starting state s_i. It is small for states of low and moderate stress and much higher for the high levels. The magnitude of the standard deviation is similar to the mean value.

Discussion The results presented above are valid for regular transition matrices. That is, transitional matrices such that after some number of steps N, it is possible to encounter any other state, no matter what the starting state is. The matrix P^N has to have no cell with 0 entry for some N.

The other types of Markov chains, absorbing and cyclic, have different properties. Since these cases are less common in psychology, we let the reader refer to other references, such as Kemeny and Snell for details.

Conclusion Markov chains represent a fundamental framework for analyzing transitional probability matrices. By computing the limit matrix and mean first passage, psychologists can understand better the long term dynamics of a system. They may be applied on single categorical variable and configurations. If Markov chains provide *numerical computations* on transition matrix, graphical techniques are required to *visualize* such transitions. They are shown in section 4.

3.5 Conclusion

Researchers start with a sequence of observations, for which they want to extract the main dynamical properties. This chapter was intended to provide them with the necessary tools. The building block of all subsequent analyses performed on categorical time series data is the matrix of transitions. They allow researchers to determine what are the general tendencies of variables. They are often – should always be – coupled with time series plots and state transition diagrams, presented in the next chapter. They may be applied on single variables as well as configurations. Markov chains analyses were also presented, in order to assess the long-term behavior of the system. Next chapters will build upon this one, higher order Markov processes being one of the many analyses to be performed (cf. section 8.5). But let's first delve into the graphical representations of dynamics...

Chapter 4

Graphical representations

"You can see a lot just by lookin'"
Yogi Berra, famous baseball player

Summary. *Graphically representing data is the first fundamental step when analyzing the dynamics of a system. Various types of time series plots are reviewed, including phase spaces for categorical variables. They allow the visual detection of trends and patterns of evolution. They open the way for multivariate categorical time series graphics we developed, named dynamical Karnaugh maps and evolustrips. Given the importance of graphics in the research process and for the configural approach we developed, basic principles of graphics theory are also presented.*

If *statistical* methods for the analysis of processes are rather recent, *graphical* methods are much older (Tufte, 1986; Wainer & Thissen, 1993). The earliest known graphics of a time series dates back in the tenth or eleventh century (figure 4.1). Monks showed the change of inclinations of the planetary orbits as a function of time. The content is somewhat confused, quantities are absent, but it is nonetheless a remarkable achievement. In 1779, a Swiss scientist, J.H. Lambert represented periodic variations in soil temperature in relation to the depth under surface. In the last century William Playfair, considered by many as the founder of modern graphical design, showed an economic time series, representing the balance of imports and exports of England. Playfair invented most of the statistical repertoire used today: the bar chart, histogram, surface chart and too famous "pie-chart" (Tufte, 1986).

More recently chaos and complex dynamical systems would not be what they are today if computer graphics had not exist (Gleick, 1987; Peitgen, Jrgens, & Saupe, 1992). The butterfly-shaped graphics of the Lorenz attractor and images of fractals widely contributed to the diffusion of chaos theory ideas. They are considered by many as beautiful, themselves very attractive, and people without solid scientific background can feel that behind the elegant picture lie profound insights.

The Lorenz attractor is a very nice example of a plot representing the evolution of the

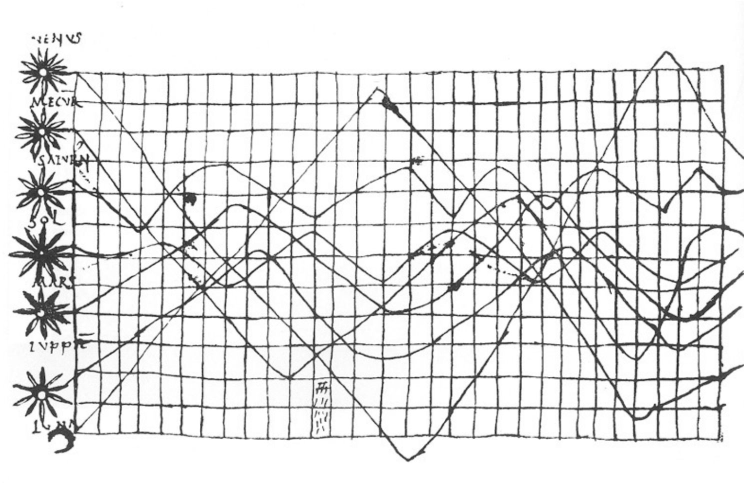

Figure 4.1: Examples of early time series graphics

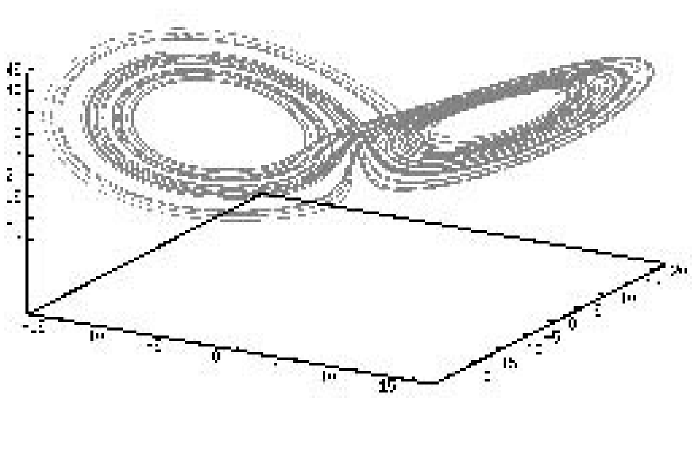

Figure 4.2: The Lorenz attractor

system across time. Mathematically it is called a *phase space* , or sometimes *phase plot* [1] It consists of two- or three-dimensional (Euclidean or Cartesian) graphics, where the axes are used to represent the possible values of the system, one axis for each variable. Then for each observation, defined by their values for the variables, a point is drawn in the phase space. A line is added to link all these points in their actual sequences. Abraham in his artfully illustrated book *Complex dynamical systems* (Abraham, 1990) shows many examples of phase space of systems. Point, cyclic, quasi-periodic and chaotic attractors are described in details.

4.1 Exploratory data analysis

There are mainly two stances in statistical analyses: exploratory and confirmatory. The exploratory stance is like a detective work: searching for signs of events, mostly through visual explorations of patterns of data. The confirmatory stance is like a judicial work, weighting evidences against competing hypotheses: statistical tests are the required tools. Both reflects two different moments in the steps towards a better comprehension of the phenomenon under scrutiny.

Even if the final objective may be to test a specific hypothesis, the exploratory phase is a primary importance (Tukey, 1977). It is an important initial step in the investigation of psychological data. "EDA is concerned primarily with explorations and description of data, not with inference. The techniques are designed to identify fundamental, conceptually meaningful patterns and relationships in data and to call attention to observations that deviate greatly from those fundamental patterns." (Smith and Prentice, 1993, p.355)

One of the main approach for data exploration is graphical. Tukey (1974, p.526) wrote that "the picturing of data allows us to be sensitive to not only to the multiple hypotheses we hold, but to the many more we have not yet thought of, regard as unlikely or think impossible." EDA recommends extensive use of graphical displays. Stem-and-leaf plots and boxplots are well-known examples of EDA displays for synchronical analysis.

As a "pre-processing" phase, screening data is important because detection of relevant characteristics influences decisions to be taken later on. With data not normally distributed or stained with outliers researchers may have to transform data or remove these too different observations. Screening data helps to choose the appropriate transformation.

4.2 Graphics theory and data displays

"What is to be sought in designs for the display of information is the clear portrayal of complexity. Not the complication of the simple; rather the task of the designer is to give visual access to the subtle and the difficult - that is, *the revelation of the complex*." (Tufte, 1986, p.191)

Through out this work we constantly make use of various types of graphs and animations. And since we even developed various graphical techniques for representing the dynamics of systems (evolustrips and dynamical Karnaugh maps), we found it necessary to

[1] To be accurate the phase space is the mathematical space of all possible values of the system. The term has come to refer to the graphical representation itself.

review the main concepts and principles of graphical theory.

4.2.1 Purposes of graphics

Graphics are employed for four purposes: depiction, exploration, summarization and communication (Wainer and Thissen, 1993, p.450). Their goal is not to add decoration and colors to a scientific work, for public relations, or to distract readers from the hard facts. They answer some important needs.

Why graphics are so useful? Graphics allow to:

1. display enormous amount of information in a small surface;
2. synthesize information in a visual manner;
3. communicate information easily;
4. explore structure of variables and their relationships: it may later help to transform data;
5. help seeing the unexpected: "The greatest value of a graph is when it forces us to see what we were not expecting" (Tukey, 1977).

4.2.2 Principles of graphics theory

"Above all else show the data" (Tufte, 1986, p.92)

Through out the use of graphical displays, we shall remember the principles of graphical excellence, enunciated by Tufte (Tufte, 1986). Most of all, people looking at graphics should not exclaim "Oh what a beautiful figure!", but "Oh, what interesting data!".

1. Graphical excellence is the well-designed presentation of interesting data – matter of *substance* , of *statistics* , and of *design* .
2. Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.
3. Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
4. Graphical excellence is nearly always multivariate.
5. Graphical excellence requires telling the truth about the data.

The advent of computer graphics has increased the visualization capability of researchers but unfortunately, has also generated a flock of badly design graphics. Ugly and uninformative graphical displays are more common than ever. Pie-charts is surely the paragon of bad data graphics, and so are the automatic inclusion by software of diverse useless backgrounds (such as moiré patterns), frames, ticks and pseudo-enhancements. Esthetics could be a concern for the communicator, but rarely do researchers possess the necessary *doigté* to do it well.

To escape the bad design avenue Tufte advocates the *data-ink ratio* as a reference. The data-ink ratio is "the proportion of graphic's ink devoted to the non-redundant display of data-information." (Tufte, 1986, p.93). He shows that even simple graphics such as histograms repeat the same information about 5 times; the notion of frequency is coded by the height of the each bar, but the height itself is expressed by: lines on the left, right and top side of the bar; the color inside; the numeric frequency at the top of the bar.

The idea is to maximize the data-ink ratio, *within reason* . Redundancy is not to be avoided at all costs. It sometimes helps the reader to access information by different modalities. It simply should not be over-represented.

4.2.3 Limitations of graphics

The disconcerting facility to make and analyze graphics has brought a tendency in some psychological circles to rely exclusively on graphical representations. Visual inspections gained wide acceptance among researchers and practioners working in schools, clinics and other applied settings (Franklin, Gorman, Beasley, & Allison, 1996).

Even if visual inspection should be encouraged as much as possible, researches have shown that even trained specialists make a few errors when visually interpreting trend lines, to name but a few major type of errors (Franklin et al., 1996). Perceptual and cognitive bias (Deschamps & Clémence, 1987) are further source of errors limitating the sole reliance on graphics eyeballing (Franklin et al., 1996).

Beside *interpretative errors* by practioners, there are also *design errors* committed by graphic makers. There is a small book entitled "How to lie with statistics" (Huff, 1986). There should be a book entitled "How to lie with graphics". It is quite easy to mislead readers with deceptive graphics, even using real figures.

Again following Tufte's principles, graphics are badly designed when they are:

- out of context graphics;
- showing disproportionate effects (especially manifest with area graphics);
- showing wrong perspective;
- missing, unclear, undetailed labels;
- showing of design variation, not data variations;

We shall remember this advice when make use of graphics for representing the evolution of systems.

4.2.4 Conclusion

In this section we overviewed the principles of graphics theory, principles that will serve as guidelines when building and interpreting data displays. We strongly advocate the visual inspection of data before performing or in conjunction with statistical analyses. Limitations of graphical representations were briefly summarized.

4.3 Graphics for time series data

4.3.1 Time series plots for continuous variables

Exploratory analysis emphasized the necessity to visually explore data before attempting to perform any numerical computations. Analyzing dynamics is no exception.

There are two techniques to graphically represent the system's dynamic: *time series plots* and *phase spaces* . The examination of time series plots gives an idea of the system *trends* , *variability* and *phases* (Franklin et al., 1996). Phase spaces reveal in a compact format *transitions* that occurred during the period of observations. They are presented in section 4.3.2.

We first review graphics for continuous time series data. They will serve as a reference when introducing graphics for categorical time series data.

Method For univariate quantitative variables, the typical graphics is the time series plot. The data point taken by the variable at time t is placed on the vertical axis, while time t is put on the horizontal axis.

Example The evolution of the full scale stress variable of our subject example is typically represented as in figure 4.3 [2].

Since readers may interpret the continuous line of the plot as a sign of a continuous data recording - which is not the case - and to better emphasize each level of the variable, we suggest following Tufte's advice to improve the traditional time series plot and make it as of figure 4.4 [3]. Each observation at time t is represented by a tick square and successive observation are connected through a vertical line (instead of an oblique one). If horizontal lines were drawn in the background, the emerging picture would resemble a musical partition...

Examining this graphics brings many relevant information. First we comment on purely static tendencies. We usually start by considering the **range and variability** of the system. What states did the system encountered? What are the extreme values, and what does it tell about the system? Here we see that this person experienced all values of the six-point scale. She mostly experienced stress levels 2 to 4; cases of more intense stress were occasional (levels 5-6), and so were moments of absence of stress (level 1).

The second feature we extract is more dynamical, the presence of **trends and patterns** . Does the system have a tendency to generally increase or decrease its level? Or does it seem to be constant across the period of observation? For our example, there is no obvious emerging increasing or decreasing trend.

The presence of **phases** , local patterns during particular periods, faintly appears from the background. The first 40-or-so observations is a series of ups and downs, ranging between 2-4, with a few 5s and 1s, but no 6. Then arrives a 6, some 3s, 4 and 5, then followed by a series of 2. After that stress decreases with many observations at 1, and then return to higher levels of stress, and so on. While visually there seems to be changes of dynamics at

[2] Produced with the S-Plus command:

```
> tsplot(STRE, main="Evolution of stress", xlab="Time (t)", ylab="Stress
intensity", axes=T, bty="n")
```

[3] Produced with the S-Plus command:

```
>  TSPlot(STRE, title="Evolution of stress")
```

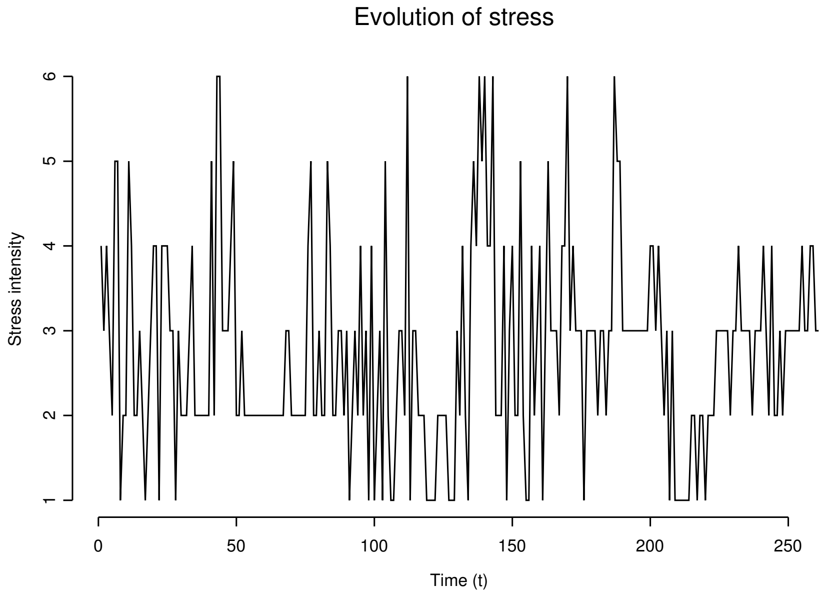

Figure 4.3: Evolution of stress variable across time

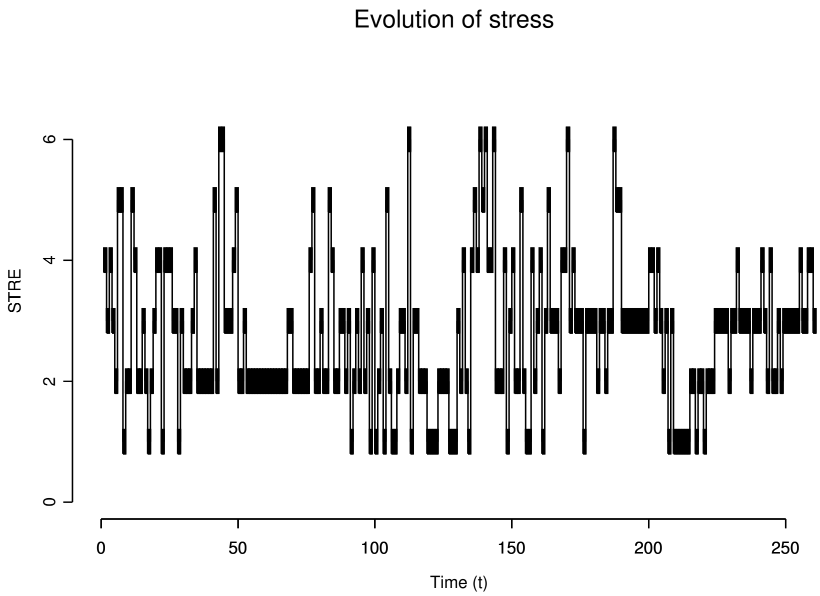

Figure 4.4: Evolution of stress variable across time

different periods, it is somehow difficult to describe verbally. Of course the type of description we produce depends on the time scale we use (a few observations at a time, or blocks of tens). Further statistical tests will be necessary to answer the question.

The detection of phases is an important issue, because most of the subsequent statistical analyses require that the data are **stationary** , a property implying that the dynamics remains stable for the investigated period (we say dynamics, not data, otherwise there would not be any dynamics; more on that in section 7.1).

Analyzing in details a graph of evolution may also indicate the presence of **periodicity** , states recurring at specific intervals. If the system exhibits a periodicity of length k, we would see a repetition of k states. In this example, no cycle seems to strikingly emerge.

Discussion Time series plots are a definite improvement over the simple examination of data values (remember the inspection of the raw data in section 3.1). They are the devoted graphical technique to provide insights about different characteristics of univariate quantitative dynamical systems: range and variability, trends and patterns, phases, stationary and the likes.

There are two major limitations of these graphics. The first one is that it is unsuitable for representing categorical time series data. Some improvements are needed in order to deal with the "non-quantitative" nature of data; this point is discussed in section 4.4. The second limitation relates to its incapacity to adequate manipulate more than one time series.

4.3.2 Phase spaces for continuous variables

Time series plots are an adequate technique for visualizing the global trend of the system. However it is sometimes difficult to figure out the underlying patterns of transitions. Moreover it is difficult to integrate the dynamics of many variables. Phase spaces answer these needs.

Method The phase space is obtained by graphically representing the state of the system not as a function of time, as do time series plots, but as a function of the last state(s).

The are two types of phase spaces: the "univariate" version, where the time axis is replaced by the state of the considered variable at lag 1 (yielding a two-dimensional graphics), or at lags 1 and 2 (yielding a three-dimensional graphics).

Examples

The univariate 2D and 3D lagged phase space On figure 4.5 [4] is represented the two-dimensional phase space of the stress variables of our subject. On the x-axis is stress at time t and on the y-axis is stress at time $t+1$. Each data point is "jittered" a little in both x and y directions, otherwise we could not distinguish transitions that are on the same line.

[4] This graphics is produced with Splus 4.5, using the Graph Menu, 2D Plot, Scatterplot.

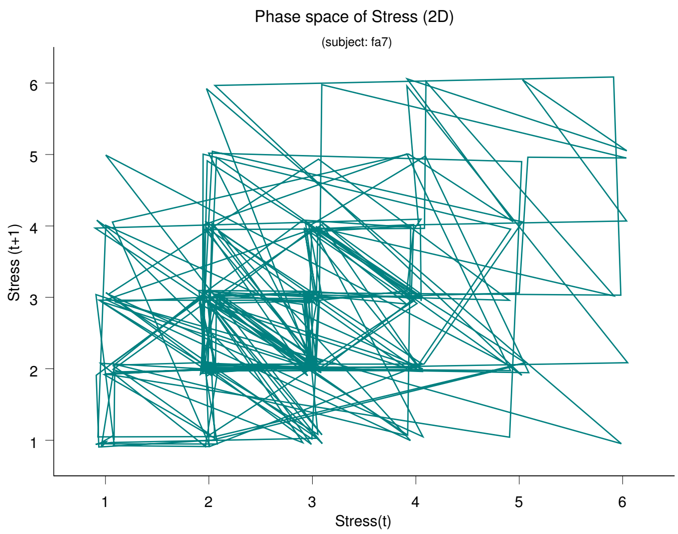

Figure 4.5: Phase space of stress variable (2D)

The univariate 3D lagged phase space on figure 4.6 [5] represents the three-dimensional phase space of the stress variable of our subject. On the x-axis is stress at time t, on the y-axis is stress at time $t+1$ and on the z-axis is stress at time $t+2$.

The resulting pictures resemble a spider's web. Transitions enter and leave almost all states, showing a "relative" disorder in this subject's dynamics. But the high density of lines in the interval STRE(t,t+1)= [2,3] informs that many transitions occurred between these states.

The lower density of lines around states 5 and 6, and to a lesser extent at 4, reveals that higher stress level did not occur frequently; the absence of transitions from 1 to 6 is also visible.

The multivariate 2D and 3D non-lagged phase space The phase space is probably more useful for considering different variables. We will see how the dynamical relations may be detected using multivariate phase spaces.

Let's start with two variables. Suppose we are interested in the co-evolution between the two internal variables stress (STRE) and emotionality (EMOT). Their two-dimensional phase space is constructed as on figure 4.7 [6]. On the x-axis is stress and on the y-axis is emotionality at time, where points refer to the bivariate state at the same t. Therefore contrary to the last case, a point in phase space refers to the state of the system *same moment* , and not a two different times.

We go one step further by introducing a third variable into our last phase space. There is

[5]This graphics is produced with Splus 4.5, using the Graph Menu, 3D Plot, 3D Scatterplot.
[6]This graphics is produced with Splus 4.5, using the Graph Menu, 2D Plot, Scatterplot.

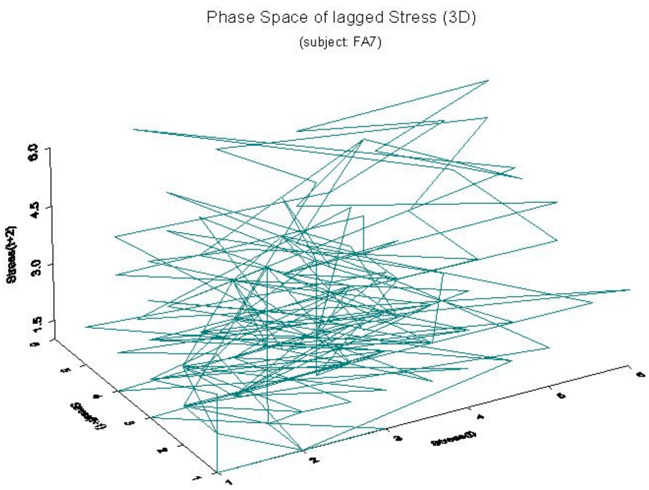

Figure 4.6: Phase space of stress variable (3D)

now a z-axis, which represents the familiarity of situation (FAM). The resulting picture is on figure 4.8 [7].

As with the univariate two-dimensional phase space the resulting pictures also resemble a spider's web. But they seem more organized than the previous examples. On a static level a strong linear relation exists between the three variables. Indeed most observations are found on a line, where both variables are equal; and there are no cases of high stress and low emotionality, and vice-versa. On a dynamics level, transitions seem to be directed towards states of low stress and emotionality.

Discussion At this stage the technique could have been imperfect, if we had not "jittered" data points. Because of the regular grid formed by the state values, we could not have differentiated transitions going from one state to another; how can we tell the difference between a transition from STRE=2 to STRE=3 or STRE=5? or one going from STRE=4 to STRE=1 or STRE=6. It would have been strictly impossible. Moreover we could not have distinguished infrequent and very frequent states. Fortunately this little "trick" allows to make these two fundamental distinctions.

Other transformations to this type of graphics are possible. Here is a list of possible relevant modifications:

[7] This graphics is produced with Splus 4.5, using the Graph Menu, 3D Plot, 3D Scatterplot.

Figure 4.7: Phase space of stress and emotionality variables (2D)

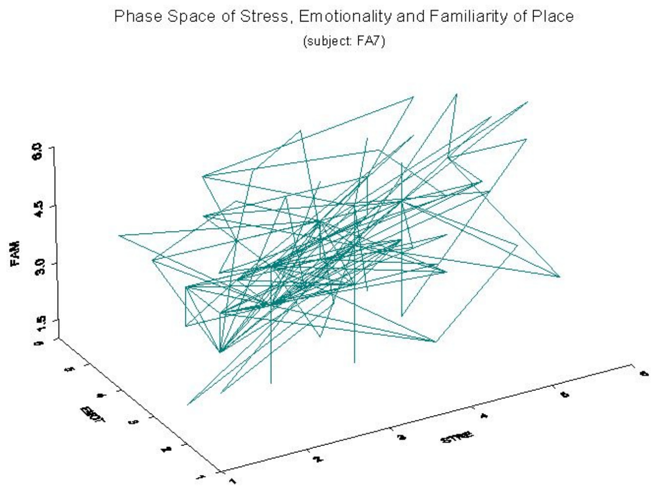

Figure 4.8: Phase space of stress, emotionality and familiarity of situations variable (3D)

1. add points varying in size according to the frequency of the transition;
2. make curved lines to represent transitions; this removes the problem of not knowing to which state the transition is directed to;
3. make arrows instead of lines, to see in what direction the transition goes to;
4. make the arrows or lines thinner or thicker according to the transitional probability for the originating state;
5. add a color code so that a third or fourth variable is taken into account.

Conclusion Interesting patterns of transitions emerge when visualizing the evolution of data. Various graphical representations were shown in this section: time series plots and phases spaces, in univariate and multivariate, two-dimensional and tri-dimensional versions. The appropriate type is selected according to the researcher's objective.

The purpose of describing graphics for continuous time series data was to set references for categorical data. Now let's examine how similar techniques may be used on categorical data.

4.4 Evolustrips: time series plots for categorical data

As with interval variables, a typical categorical time series plot should put time on the x-axis and the dependent variable on the y-axis. However the direct utilization of continuous variable time series plots causes drawbacks, because there is no definite order in the categories of nominal variables. Consequently how should the y-axis be organized? With no particular order one can only tell if the variable state has changed or remained the same. A drop of two levels on the graphics is not a change of two units in the variable, but only a change of level. Furthermore the choice of the value order may influence the perception of the dynamics of the system.

An example of a time series plot for a categorical variable, PLACE, is in figure 4.9 [8]. Because of the graphic format anyone would easily interpret changes in the line as variations of the *intensity* of place, but there is no such intensity variations. The graphic is also highly dependent on the codes for categories: regular places were coded as numbers from 1 to 4 but the place "non-specified" was attributed code #9; see what consequences this choice has on the graphic outlook. A more appropriate graphical representation is required.

4.4.1 Univariate evolustrips

We now introduce a variation of the traditional time series plot for dealing with categorical variables. It adequately answers its purpose of representing the dynamics of a system. Instead using the *height* on the y-axis to represent categories of a variable, we use *color* . Each category of a variable is assigned a specific color. As time increases on the x-axis, the state of

[8] Produced with the S-Plus command:

```
> tsplot(PLACE, main="Time series plot of PLACE (Subject:  FG9)",
xlab="Time (t)", ylab="PLACE", axes=F, bty="n")
```

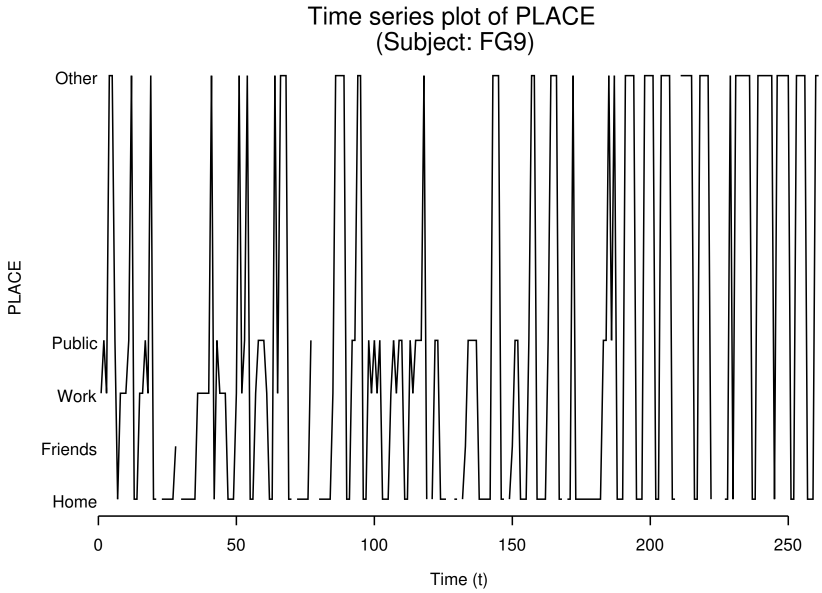

Figure 4.9: Time series plot of categorical PLACE variable

the variable at time t is given by the color on the band. We nicknamed this graphical device the "evolustrip".

Figures 4.10 [9] and 4.11 represent the evolution of the PLACE variable for two subjects on an evolustrip. The first encountered category is assigned the light blue color, the second one is pink, the third is light green, then gold, medium blue, and so on (these colors correspond to the default S-Plus color order; but they could easily be changed with a single command).

The dynamics of the PLACE variable for these subjects are quite different from each other. It unambiguously shows that the first subject stayed at home for a period of over 100 days in the second half of the experiment. The first half is mainly characterized by an alternation of work, public places and friends' place. The drastic change in the dynamics is explained by an accident of the subject, leading her to stay home for a long period of time. The dynamics is thus truly non-stationary.

For the second subject the evolustrip (figure 4.11) [10] shows that the evolution might also be a non-stationary process. Scattered throughout the evolustrip is the green color, the code for the "At home" place, which is quite normal. This is some concentration of black ("At work") during the first third, light blue ("In public places") during the second third and pink ("Other places") during the last third. This might be explained by the period of observations, set in spring time (1994), which crossed a session of examinations for the students, followed

[9] Produced with the S-Plus command:

```
>  Evolustrip(PLACE, title="Evolustrip of PLACE (Subject:  FA7)")
```

[10] Produced with the S-Plus command:

```
>  Evolustrip(PLACE, title="Evolustrip of PLACE (Subject:  FG9)")
```

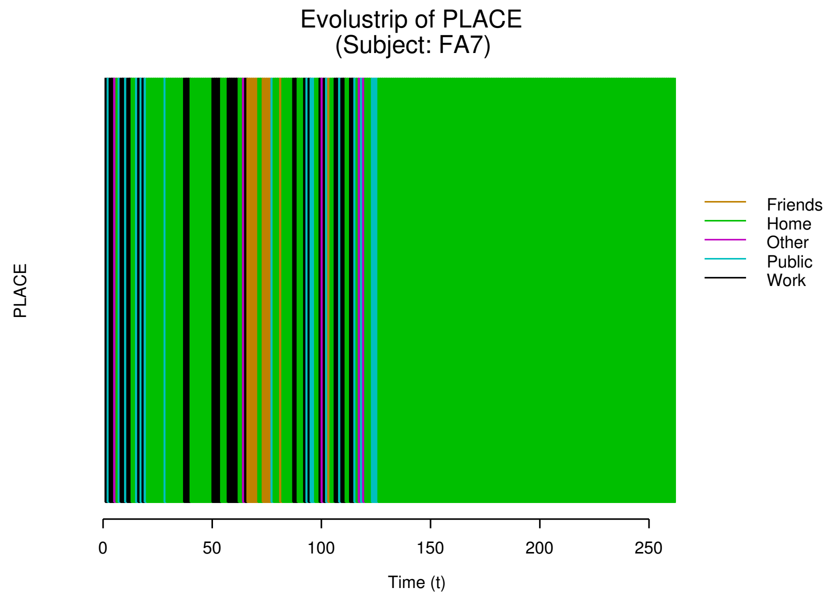

Figure 4.10: Evolustrip of PLACE variable (subject: FA7)

by an holiday break from university.

4.4.2 Multivariate evolustrips

It was previously shown that employing a traditional time series plot for displaying the parallel evolution of many categorical variables would not be an advised choice. At the same level on the y-axis correspond different categories for the various variables. The superposition of lines at the same level confuses the reader.

The use of the evolustrip technique makes feasible the extension to numerous categorical variables, thus making it a *multivariate evolustrip* .

Method To transform the univariate evolustrip into a multivariate one is simply a matter of adding strips to the graphics, one for each new variable. The number of strips in the graphics then tell how many variables are employed. The use of colors is still the method for differentiating categories of variables.

Examples In the following example (figure 4.12) [11] is shown the multivariate evolustrip of 7 variables of our subject example: MOOD (mood level), EMOT (emotionality), STRE (stress), PRESS (external pressure), MATF (material support), SOCS (social support)

[11] Produced with the S-Plus command:

```
>  MultiEvolustrip(FA7, title="Multivariate Evolustrip of 7 Variables")
```

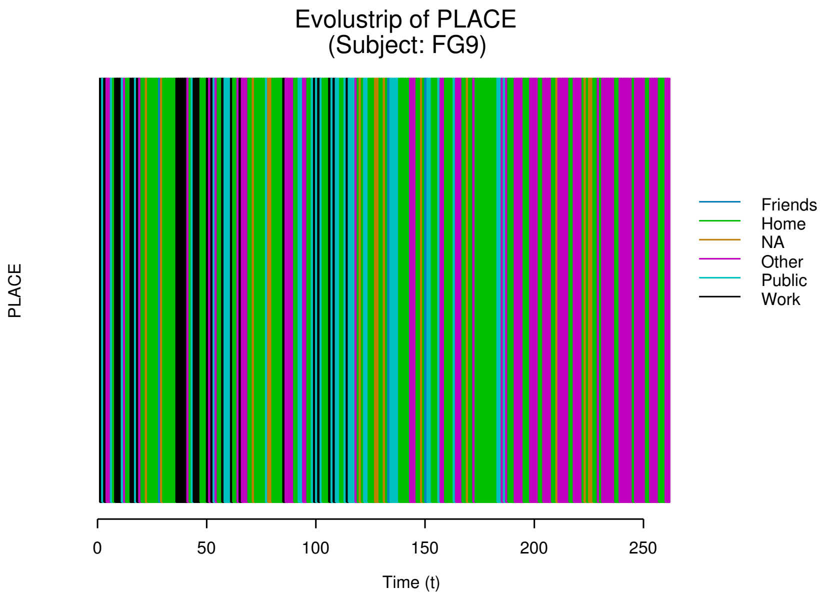

Figure 4.11: Evolustrip of PLACE variable (subject: FG9)

Figure 4.12: Multivariate evolustrip of categorical variables

and FAM (familiarity of situations). They are all ordinal variables measured on a six-point scale; the 0 value is attributed to variable "Social Support" (SOCS) when the person was alone.

The evolustrip is well suited for binary variables. As we later describe the configural phase spaces based on dichotomized variables, the Karnaugh map, we present here the evolustrip of dichotomized variables (figure 4.13) [12].

The evolustrip allows not only to plot the evolution of many variables for one subject, but also the same variable for many subjects. A researcher may then compare the dynamics of various subjects by visually inspecting the graphics.

[12]Produced with the S-Plus command:

```
> MultiEvolustrip(BFA7, title="Multivariate Evolustrip of 7 Binary
Variables")
```

Figure 4.13: Multivariate evolustrip of binary variables

Conclusion The evolustrip is a well suited graphical technique for representing and exploring the temporal evolution of categorical time series data. Like its quantitative counterpart, the time series plot, it allows researchers to visually explore the structure of the dynamics, especially the general trend, variability and spot the presence of phases. Its main advantage relates to its capacity to plot multivariate dynamics.

4.5 State transition diagrams

In section 4.3.2 were presented phase spaces, a technique for plotting quantitative variables patterns of transition. Using the same type of graphics would be interesting for categorical variables but in a different format. Here we present the appropriate graphical technique for representing the dynamics of categorical variables: the *state transition diagram* .

Method State transition diagrams synthetically represent with arrows how transitions occurred, from a state i an other j. As with phase space, there is no time axis, but instead a "folding" of space so to plot only the names of states.

How are state transition diagrams drawn? The name of each encountered state (category) is written, and then arrows are added for each transition observed, going from state i to state j. An example will clarify this matter.

Example We present on figure 4.14 [13] the state transition diagram for the place variable. Direction in transitions from state i to j is to be interpreted in a clockwise fashion; lines at the top represent transitions from states on the left side to states on the right side while lines at the bottom are for transitions from states on the right side to states on the left side. A circle around a state indicates a transition towards itself. The numbers indicate the percent of transitions from state i to j.

This state transition diagram graphically summarize how transitions occurred in the course of observation. But synthesizing information is a difficult task because of the high number of lines.

Most frequent transitions When considering variables with more than two or three categories the diagram becomes confusing because of the high number of represented links. As a first transformation we suggest to plot only the most frequent transitions. Only the transitions that have the greatest frequency for each state i are plotted (thus there are only m transition plotted, where m is the number of different states). Later we will statistically test transitions and then plot only the significant ones.

The resulting state transition diagram for the place variable yields the picture on figure 4.15 [14].

[13] Produced with the S-Plus command:

```
>  TransitionDiagram(PLACE,Title="State transition diagram for PLACE")
```

[14] Produced with the S-Plus command:

```
>  TransitionDiagram(PLACE,Title="State transition diagram for
PLACE",mft=T)
```

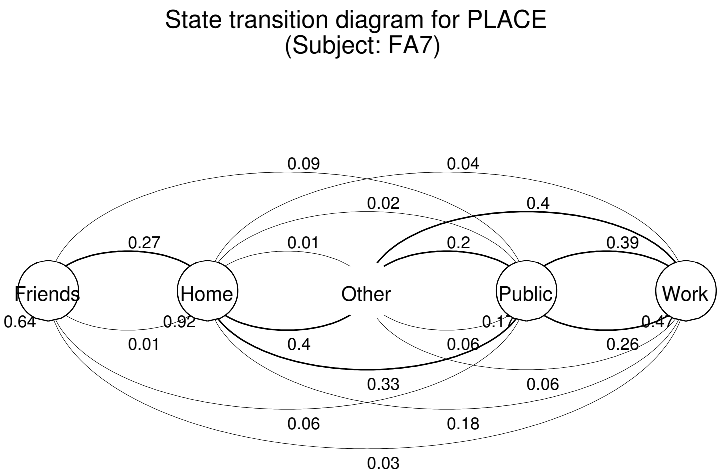

Figure 4.14: State transition diagram for place variable

The subject had the tendency to remain in the same place when she was at home (in 92% of cases), at her friends' place (in 64% of the cases) and at work (in 47% of cases). Those "states" are considered as much more stable than the others. The category "Other" is a rather unstable state, because she was as likely to return home as to go to work (in 40% of cases). The public places are most often followed by the work place (in 39% of cases).

To further exemplify our point, let's plot the state transition diagram of the three dichotomized variables shown in the configuration chapter: familiarity of situations (BFAM), emotionality (BEMOT) and stress emotionality (BSTRE). On figures 4.16 are drawn the state transition diagrams of the most probable transitions for these three variables, and finish with the configuration made up of these three binary variables.

Regarding the dichotomous variables, we can infer that the subject tried to be in familiar situations, maintain a low emotionality and low level of stress. Indeed for the BFAM variable, the most frequent transitions are towards the BFAM=1 state, while for the two internal variables the MFTs are towards the BEMOT=0 and BSTRE=0.

When considering the MFTs for the configuration composed of these three variables, it becomes obvious that the configural state #4 (familiar situations, low emotionality and low level of stress) acts like an "attractor". All MFTs except #6 are directed towards this configural state. So whatever situations she encounters she is most likely to be found in this pivotal state.

Although certain patterns of transition emerge from the picture, Karnaugh map yield a better representation of the underlying dynamics.

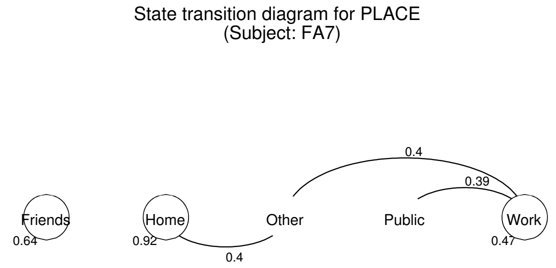

Figure 4.15: State transition diagram for MFT of PLACE variable

Discussion We suggest using the state transition diagram technique for understanding the fundamental univariate dynamics of our subjects. Plotting these state transition diagrams for each variable of each subject of our study has shown that most of the subject tried to maintain low levels of stress and emotionality, find social and material support in their environment, be in familiar situations, and so on. This is on a dynamical basis: if they were in a state of high stress, they would try – and succeed – to return to a state of low stress. This is shown simply by plotting the most frequent transitions for each state.

One limitation of this technique is that drawing transitions for a number of states greater than two or three makes it difficult to understand, the state transition diagram becoming cluttered by too many arrows. Plotting only the most frequent transition from each state somehow alleviate the problem but not completely.

Conclusion The state transition diagram is an adequate graphical technique for representing the transitions between categorical states. Investigators see right away what states were encountered and how transitions from each state occurred. Beside links between states, frequencies or probabilities may also be written down. It adequately complements the evolustrip.

Even if considering variables with more than 6 categories brings plotting problems there is a surprisingly neat solution for configurations of up to 16 or 32 states, if they result from groups of binary variables. The technique is called dynamical Karnaugh map, and is presented in the next section (4.5.1).

4.5.1 Dynamical Karnaugh maps

We have shown previously how to represent the synchronical organization of configurations, with Karnaugh maps (cf. section 2.3). We will see how to combine the structural properties of this graphics with the dynamical features of state transition diagrams: the result is the

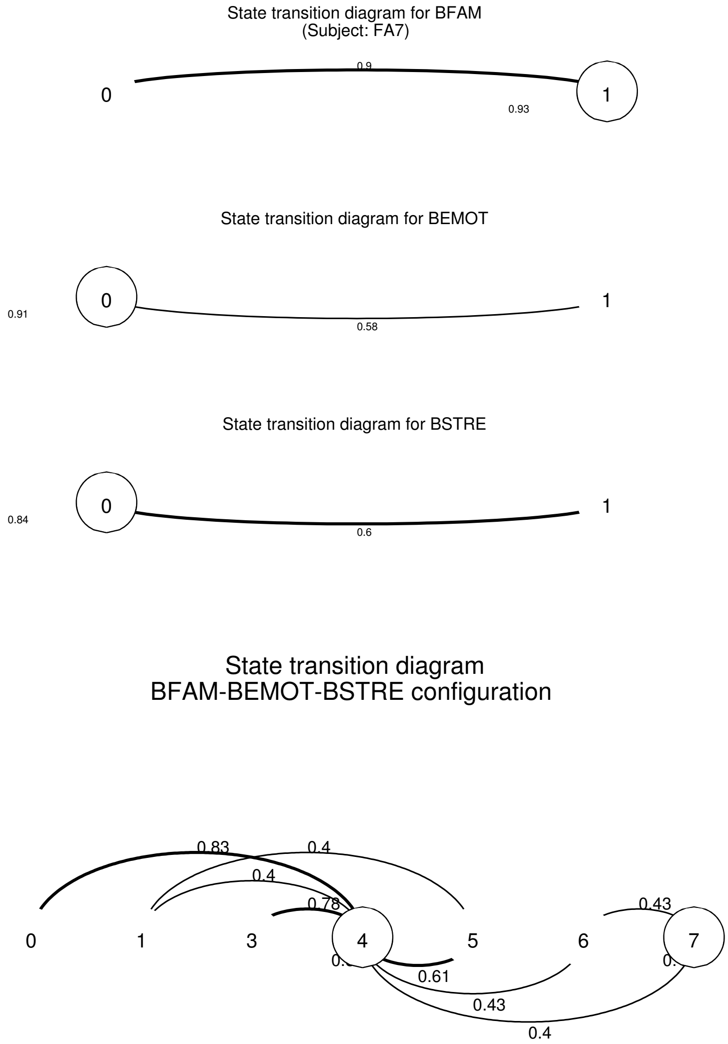

Figure 4.16: State transition diagrams for the MFT of dichotomous familiarity of situations, emotionality and stress variables

dynamical Karnaugh map. It is a truly multivariate state transition diagram, or stated more simply, a configural transition diagram.

Like in any phase space, the successive observed states are connected by a line (from the state at time t to state at time $t + 1$). As will be shown later the evolution of a complex system is easily plotted and tracked, as its bifurcations and phase transitions appear in an integrated and economical fashion. Last but not least, this tool has shown definite marks of self-organizational processes for the subjects participating in our study.

Method At this point we merge two very fundamental graphical techniques we have seen so far: the state transition diagram with the structural Karnaugh map. It yields the dynamical Karnaugh map.

The principle is to construct a structural Karnaugh map with a set of 3 or 4 variables, and then connect each successive configuration with arrows according to the observed transitions from each configural state (as obtained by the transitional frequency matrix, explained in section 3.2).

Example As an example, let's examine the Karnaugh map for our subject. The variables chosen are BFAM, BEMOT and BSTRE. It is shown on figure 4.17 [15].

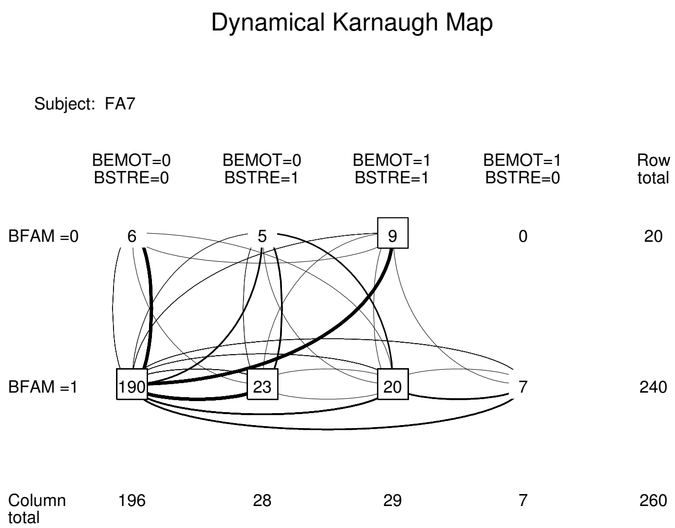

Figure 4.17: Karnaugh map for BFAM, BEMOT and BSTRE variables

At first the visualization of such figure leads more to confusion than understanding. We

[15]Produced with the S-Plus command:

```
>  Karnaugh3(BFAM,BEMOT,BSTRE,subjectname="FA7")
```

mainly see that all configuration states were encountered, except one (configuration #010, $BFAM = 0$, $BEMOT = 1$ and $STRE = 0$). More interesting, transitions seem to appear from each 7 configural states towards all other 7. This implies a diversified period for this subject.

Plotting all configuration transitions is probably a waste of time – and ink. In almost all cases for which we plotted such graphic, a blurred picture appears. There is a better way to synthesize the information about the dynamics without overwhelming or confusing investigators and readers of such map.

A simple, efficient proposition is to consider only the MFTs, the most frequent transitions. As explained before, instead of taking into account all transitions from each configuration state, only the transition with the highest probability is retained and plotted. This solution allows to retain the most "predictive" state given each state. It represents what the system is most likely to be found if we know in which state it is in. Figure 4.18 [16] represents the MFTs of the last Karnaugh map.

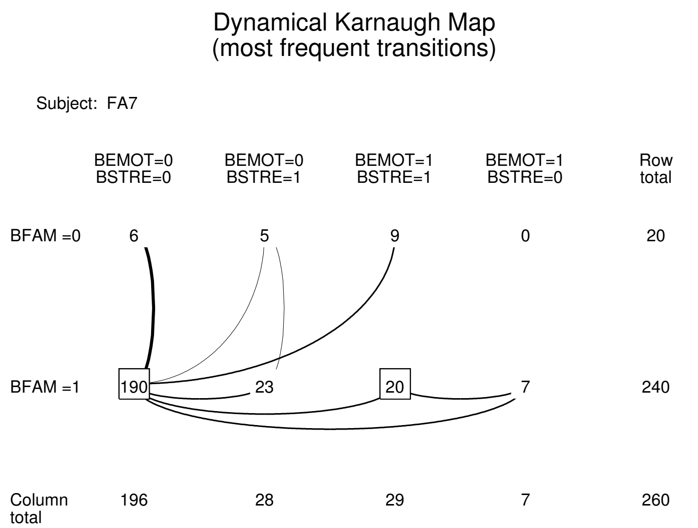

Figure 4.18: Karnaugh map of MFTs for BFAM, BEMOT and BSTRE variables

From this simplified Karnaugh map emerges a much more interesting picture. The apparent fuzziness of the dynamics was produced by close to irrelevant transitions. There is an order, an organization behind the sequence of observations. The Karnaugh map of most frequent transitions reveals that this subject spent most of her time in a "comfortable" configuration, #100, exhibiting a familiar situation (BFAM=1), low emotionality (BEMOT=0) and

[16]Produced with the S-Plus command:

```
>  Karnaugh3(BFAM,BEMOT,BSTRE,subjectname="FA7",mft=T)
```

low stress (BSTRE=0). Considering transitions from each configural state, we see that although she encountered other kinds of experiences than this easy #100 configuration, she (almost) always would return to that configural state the next moment (2 hours later, the interval between each self-observation).

But there are two exceptions to this "dynamics rule": from state #010 she mostly went through state #110, implying she would first try to find a familiar situation, before reducing her stress (indicating an external attributional style?); the other exception is for state #111, from which she equally stayed in that same state than transited to the #100, easy configuration.

Considering the description given above configuration #100 may well receive a particular name: an *attractor* . Complex dynamical systems theory describes an attractor "a state towards which a system returns to, even after some perturbations" (Haken, 1983). The metaphor of a ball in a bowl is illustrative of the type of attractor we encounter in our data. Throw the ball in the bowl, it will roll and move around for a while, and sooner or later it will gradually slow down and immobilize at the bottom. Shake the bowl a little, the ball will move but return to the bottom. The bottom of the bowl is the attractor of the system.

The same happens here with our subject. The attractor configuration is the familiar situation with low stress and emotionality. Bring some (external) perturbations, source of potential stress and emotions, the person will momentarily experience something else, but through diverse coping strategies she will quickly return to this central configural state.

As an example of a quadrivariate dynamical Karnaugh map, a fourth variable is added to the preceding graphics: PRESS. The dynamical Karnaugh map of the most frequent transitions is shown on figure 4.19 [17]. The resulting picture makes it clear that the configuration #1000, characterized by a familiar situation (BFAM=1), low environmental pressure (BPRESS=0), low emotionality (BEMOT=0) and low stress (BSTRE=0). The subject really tended to stay in this configuration or return to it if changes occurred.

This same kind of affairs is found for all of our subjects; they all return to this same configuration #1000. As an example, we show on figures 4.20 [18] and 4.21 [19] the dynamical Karnaugh map of the most frequent transitions for two subjects. A few transitions do not go directly to the #1000 attractor, but they occurred from very infrequent (n=2) configurations.

A last example will show a Karnaugh map of a random process (for each of the four variables 261 values were generated by a binomial random process of mean 0.5). The resulting Karnaugh map of MFTs is shown on figure 4.22 [21]. As expected no particular patterns of transition emerges from the map. It is surely not a proof that the other subjects are truly self-organized, by it nevertheless constitutes an interesting comparison.

[17] Produced with the S-Plus command:

```
> Karnaugh4(BFAM,BPRESS,BEMOT,BSTRE,subjectname="FA7", mft=T)
```

[18] Produced with the S-Plus command:

```
> Karnaugh4(BFAM,BPRESS,BEMOT,BSTRE,subjectname="FG9", mft=T)
```

[19] Produced with the S-Plus command:

```
> Karnaugh4(BFAM,BPRESS,BEMOT,BSTRE,subjectname="XR3", mft=T)
```

[21] Produced with the S-Plus command:

```
> Karnaugh4(BFAM,BPRESS,BEMOT,BSTRE,subjectname="Random", mft=T)
```

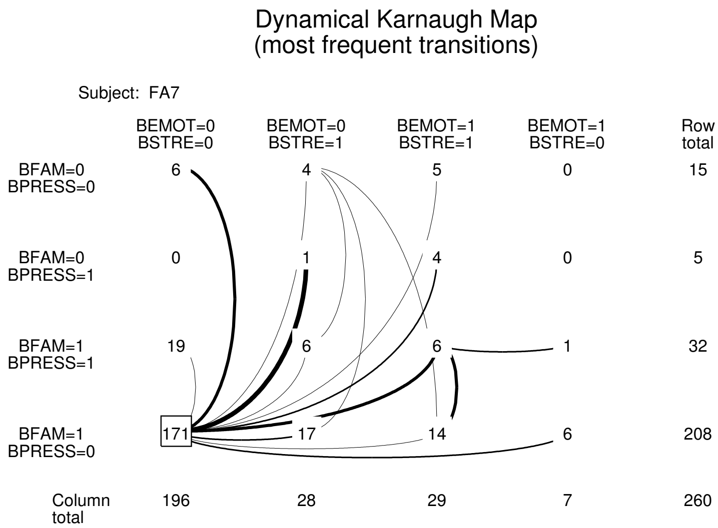

Figure 4.19: Karnaugh map for BFAM, BPRESS, BEMOT and BSTRE variables (subject: FA7)

Discussion The dynamical Karnaugh map is a very interesting technique for plotting the synthesized evolution of multivariate configurations. Its benefits are manifold:

1. allows the consideration of many categorical variables;
2. allows a *micro-macro* perspective, considering both individual variables and whole configurations;
3. shows which configuration(s) were went through, the most and least frequent;
4. shows around which configuration(s) the system revolves (the emerging "attractor");
5. show a possible adaptation mechanism (reduction of stress, emotionality, ...).

To render an even more informative graphic, arrows could have their width (or color) varying as a function of the transitional probabilities; it would then be easier to differentiate between more and less frequent transitions.

Its main limit concerns that with more than 5 or 6 variables it becomes a cluttered graphics (but any graphics becomes cluttered with this number of variables). Another limit relates to the absence of suitable representation for variables other than binary. The Karnaugh map is a very interesting graphics because of the particular internal order of variable values (the Gray code order), yielding the adjacent cell property. We did not find yet a similar order for non-dichotomous variables.

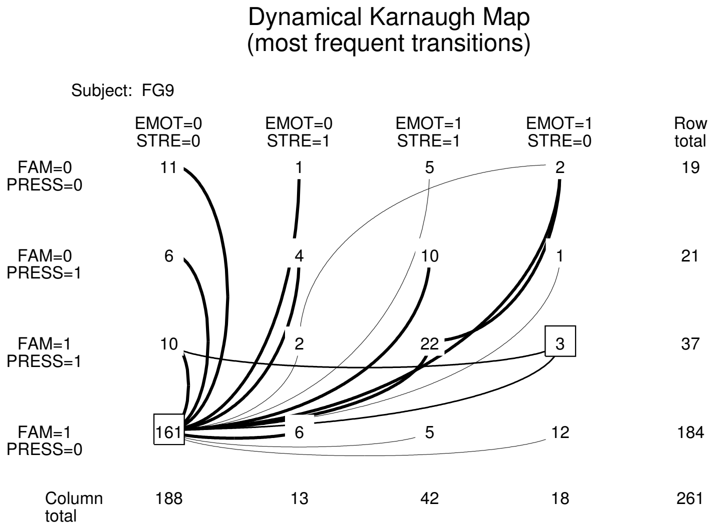

Figure 4.20: Karnaugh map for BFAM, BPRESS, BEMOT and BSTRE variables (subject: FG9)

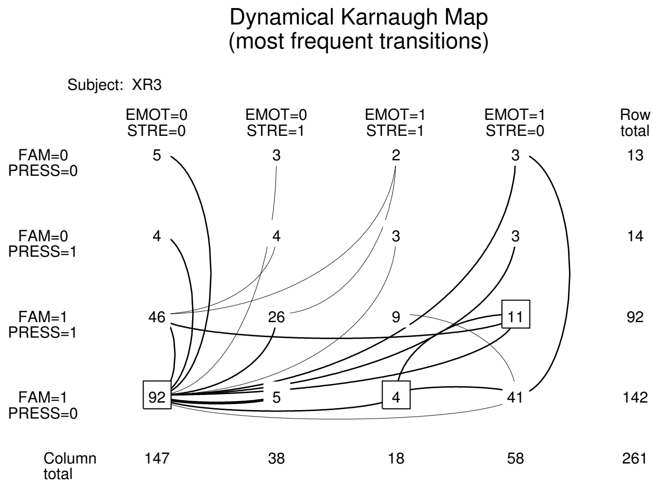

Figure 4.21: Karnaugh map for BFAM, BPRESS, BEMOT and BSTRE variables (subject: XR3)

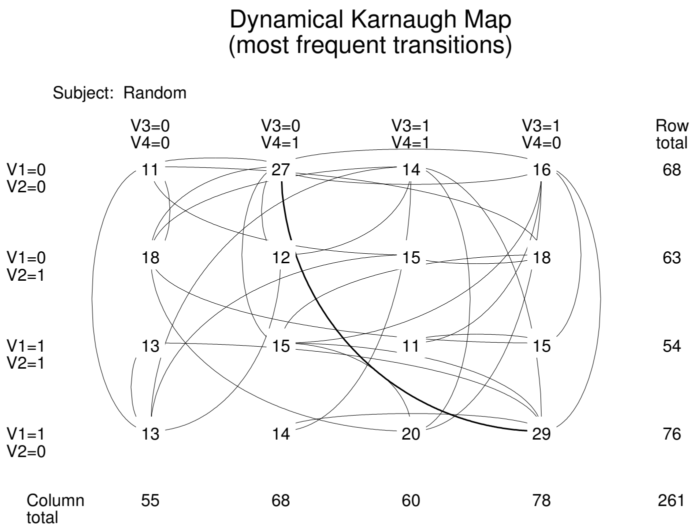

Figure 4.22: Karnaugh map for BFAM, BPRESS, BEMOT and BSTRE variables (subject: Random)

Conclusion The Karnaugh map is the graphics of choice for representing the dynamics of a system from a configural perspective. Behaving like multivariate state transition diagram it allows researchers to visualize the underlying patterns of transitions of configurations. We have shown how to draw this graphics, how to interpret it and how it yields hint about individuals' possible mechanisms of adaptation.

4.5.2 Animated Karnaugh maps

Karnaugh map, like any state transition diagram, is a "summarized snapshot" of the system dynamics. Investigators may be interested in a picture displaying the evolution of observations across time. Two possibilities are offered: one is to simply build Karnaugh maps for different sub-sequences of the whole sequence; it is then useful complement to statistical tests that would assess the presence of *phases* (cf. chapter 7). The second possibility is to build a video animation of the whole sequence. It is presented in this section.

Method Building a video animation of a Karnaugh map technically involves a series of steps summarized as follows:

1. build a Karnaugh map showing k transitions ($k \geq 2$ for showing a trail)
2. export the graphics into an external file (in GIF graphical format, for example)
3. repeat the process for each of the $n - k$ possible transitions

4. import all graphical files into a multimedia program that can build video animation from separated images
5. save the sequence of images into a video format (Quicktime, for example)
6. play the video and observe the evolution

Fortunately we programmed a routine in S-Plus that alleviate the burden of the first 3 steps. In a single command [22] investigators produce the sequence of Karnaugh maps and their corresponding graphical images. On a Macintosh, we used the iView multimedia software (from iView multimedia Limited) to integrate the single Karnaugh maps graphics into a movie; this software allows to export graphical files of a directory into a QuickTime movie. The last step is then to play the movie sequence and observe how transitions occurred.

Example We present the video animation of the evolution of configurations for our subject example. The chosen variables are the three usual ones: BFAM, BEMOT and BSTRE. Each image (or frame in video terms) represents two transitions.

The video animation is of course impossible to render on paper. Simply point your web browser to the page (where a QuickTime plug-in is required):
`http://tecfa.unige.ch/~lemay/thesis/animation.html`

Playing the complete animation shows the prominence of the #100 configuration (familiar situation, low emotionality and low stress). The first 20 observations are unstable, alternating between configurations #011, #100, #110 and #111 (mostly familiar but emotional situations). It then stabilizes around the attractor #100 (other configurations are nevertheless still present). During the period between observations 135 and 175 the subject experienced other types of situations, summarized by configurations #001 (unfamiliar stressful situation) and #101 (familiar stressful situation) and #111 (familiar emotional and stressful situation). The following days were much typical, the attractor practically being the only experienced configuration(there were only 1 unfamiliar situation afterwards).

Discussion The evolution of the sequence of configurations may be explored by the method of video animation. Contrary to the displaying of unique Karnaugh map which represents a summarized "snapshot" of the overall transitions, the video animation allows to examine the real sequence of observations, as they gradually occurred through time. Investigators may play the video at different speed, select subsequences and then relate the result with other sources of information (such as textual description of events), thereby pointing to explanation of the dynamics.

The method gracefully complements the Karnaugh map and the evolustrip. The latter allows the representation of the evolution of configurations; however *transitions* are rather difficult to extract, for which the video animation is well-suited.

The main limitation of this method concerns the potential difficulty from the observer to extract the salient features - if any stands out. If the dynamics is a combination of a definite pattern and some "noise", the pattern may not emerge clearly to the observer. Repeated

[22] KarnaughAnimation3 or KarnaughAnimation4, depending on the number of variables

visualization of the animation may be required to make sense of the dynamics. But as an exploratory tool it serves its purpose well.

4.6 Conclusion

Graphically exploring a sequence of data being of utmost importance, as emphasized by exploratory data analysis, we presented in this chapter various techniques for plotting the evolution of a system. Transposing techniques for continuous data that are time series plot and phase space, we developed innovative technique that specifically address the problem of categorical data, namely, the absence of quantified levels. The evolustrip, state transition diagram and dynamical Karnaugh map were shown to be appropriate methods for displaying the dynamics of system, should it be on a univariate or multivariate level. An animated version of the Karnaugh map was even suggested, so to better visualize the timely passage of configurations.

Of course, graphics are an essential step towards the understanding of the dynamics of a system. However researchers should not be satisfied with visual tools alone. A supplementary further step is required in order to test hypotheses or challenge intuitions that emerged from the visual inspection. The next chapter will provide sound statistical tools for more precise description of processes.

Chapter 5

Analyzing patterns of transitions

Summary. *One of the main questions an investigator asks about a dynamics is whether or not there is a particular organization or structure behind the sequence of data. Perhaps the sequence is just the result of a random process. How can investigators evaluate it? If there is a dynamics, to what degree can one make predictions? How reliable are they? What statistical tools are available to test such assumptions? This section is intended to provide meaningful answers.*

Researchers have suggested various procedures for describing and testing temporal dependence in sequences. These are performed through some measures of association in transitional matrices. There are two categories of measures of association, applied either on the global transitional frequency matrix or on a subset, thus detailing a particular pattern of transition:

- *descriptive measures*
- *inferential measures*

The first category gathers measures that are not statistical tests *per se* . They numerically relate how observations at a given time t are associated with past observations (at time $t-k$).

The second category gathers true statistical tests, typically based on a χ^2 statistics. Inferences are then allowed.

These two categories are not truly exclusive. Some measures of association are mainly used in a descriptive fashion, such as correlations and lambda coefficient, but intervals of confidence and significance tests may be computed so to allow inferences. And conversely some descriptive measures of association are computed using the X^2 statistic.

5.1 Testing global transitional dependences

5.1.1 Auto-correlation coefficient of binary variables

When dealing with *quantitative* time series data, most of the analyses are based on the (Pearson) autocorrelation coefficient. It allows to quantify the degree of linear relationship between a variable at time t and itself at time $t + k$.

When dealing with *categorical* variables one can not use such statistics. However, there are two alternatives if data are either binary or ordinal (but the latter case will not be discussed).

Through this work we emphasize analyses performed on binary variables, the building blocks of configurations. Although there is no correlation coefficient available for configurations, one may be interested in computing a correlation coefficient on individual binary variables.

Method When dealing with binary time series, researchers may revert to the traditional Pearson coefficient of correlation. But because of the corresponding structure of the 2 × 2 transitional frequency matrix, a particular formula (equation 5.1) for the autocorrelation coefficient is available (Everitt, 1992). It helps to understand with more acuity the underlying process:

$$r = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1+} * n_{2+} * n_{+1} * n_{+2}}} \tag{5.1}$$

As for the Pearson correlation coefficient, the coefficient ranges from -1 to +1. A positive coefficient, resulting from $n_{11} * n_{22}$ greater than $n_{12} * n_{21}$, means a tendency to remain in the same state (n_{11} and n_{22} indicate the system remains in the same state, while n_{12} and n_{21} indicate a change of state). Conversely a negative coefficient indicates a tendency to change state, thus to oscillate. Unfortunately, this interpretation is not completely true, since a very large frequency n_{ij} may "contaminate" the coefficient.

When the numerator is taken in absolute value, it is equivalent to the *phi coefficient* , r_ϕ (Siegel & Castellan Jr, 1988). Unlike the correlation coefficient, the phi coefficient ranges from 0 to 1 only.

The significance of the correlation coefficient is given by the Pearson correlation test; the significance may be used to test whether the auto-correlation coefficient is different from 0.

Examples Let's examine some examples using this coefficient. The first example is about a transitional frequency matrix (see table 5.1) [1] where there is a tendency to remain in the same state; the product of $n_{11} * n_{22}$ is greater than $n_{12} * n_{21}$, yielding a positive coefficient of $r = 0.46$.

	$V1(t+1)$	
	0	1
$V1(t)$ = 0	50	25
1	10	40

Table 5.1: Transitional frequency matrix yielding a positive auto-correlation

[1] Produced with the S-Plus command:

```
>  BinaryAutocorrelation(matrix(c(50,25,10,40),ncol=2,byrow=T))
```

If we reverse the frequencies so to exchange n_{11} with n_{22}, and n_{12} with n_{21}, it gives table 5.2 [2]. The auto-correlation coefficient becomes $r = -0.46$, indicating a tendency to change state most of the time, to oscillate.

	$V1(t+1)$ 0	1
$V1(t) = 0$	25	50
1	40	10

Table 5.2: Transitional frequency matrix yielding a negative auto-correlation

The auto-correlation coefficient is applied on the dichotomized stress variable. Its transitional frequency matrix is given in table 5.3 [3]. The resulting coefficient is $r = 0.24$, indicating a tendency to remain in the same state. This straightforward interpretation is not completely accurate, as a detailed analysis of the matrix reveals. The frequency of transition $BSTRE_{1\to1}$ ($n_{11} = 170$) is much higher than the other frequencies, distorting the interpretation of the computation. The tendency of this person is not to remain in the same state as previously: when stress is high (BSTRE=1), she will reduce her stress level most of the time (n_{22} is smaller than n_{21}).

	$BSTRE(t+1)$ 0	1
$BSTRE(t) = 0$	170	33
1	34	23

Table 5.3: Transitional frequency matrix of stress variable

This auto-correlation coefficient may be tested against a zero coefficient. One simply uses the correlation test on the lagged sequence. For the previous example, it shows a significant departure from $r = 0.00$; the statistics are: $t = 4.0148$, $df = 258$, $p - value = 0.0001$.

Discussion We described in this section an auto-correlation coefficient for binary time series data. It is equivalent to the Pearson correlation coefficient, for which a particular formula exists when applied to a $2 * 2$ transitional frequency matrix. It ranges from -1 to +1. It has however a different interpretation than Pearson's coefficient. A positive coefficient implies a tendency to remain in the same state, and a negative coefficient, a tendency to alternate state. It has been shown that a frequency higher than the others may distort interpretation of results. Its significance may be tested (against a zero coefficient) using the standard correlation test.

[2]Produced with the S-Plus command:

```
>  BinaryAutocorrelation(matrix(c(25,50,40,10),ncol=2,byrow=T))
```

[3]Produced with the S-Plus command:

```
>  TFM(BSTRE)
```

Conclusion After having computed a $k = 1$ auto-correlation coefficient, one may be interested in computing it for higher orders. Section 8.2 gives more details.

5.1.2 Goodman and Kruskal's lambda

Goodman and Kruskal described several measures of association between two variables (Castellan, 1979; Siegel & Castellan Jr, 1988). The question behind their investigation is the following: "How much does the knowledge of the classification of one variable improves the prediction of the classification of the other variable?" Transposed in terms of time series analysis it may be reformulated this way: "How much does the knowledge of the state of the system at time t improves the prediction of state of the system at time $t + 1$?"

Suppose a researcher wants to predict the next state of a system. Without the transitional matrix information, the best guest is to predict the state that has the highest probability of occurrence; or more precisely, the state having the largest marginal total, $max(n_{1+}, n_{2+}, n_{m+})$. But when one knows the state at time t, the best guess to predict the state with the highest probability *given* this knowledge, taking into account the transitional frequencies. It thus reduces the probability of error.

Method The general form of the index may be written as (equation 5.2):

$$\lambda_B = \frac{P[error] - P[error|x(t-1)]}{P[error]} \tag{5.2}$$

Since the true probabilities are not known they must be estimated. We then use the statistic L_B. This statistic is then computed on a transitional frequency matrix as follows:

$$L_B = \frac{\sum_{i=1}^{m} n_{M_i} - max(C_j)}{N - max(C_j)} \tag{5.3}$$

where n_{M_i} is the largest frequency in the i-the row and $max(C_j)$ is the largest column total.

The lambda statistic ranges from 0 to 1, where 0 means the antecedent does not help at all in predicting the consequent, and 1 implies a perfect prediction. In the latter case, there is one and only state j that follows a state i.

The variance of this statistic is given by formula 5.4:

$$L_B = \frac{(N - \sum_{i=1}^{m} n_{M_i})(\sum_{i=1}^{m} n_{M_i} + max(C_j) - 2\sum_{i=1}^{\prime m} n_{M_i})}{(N - max(C_j))^3} \tag{5.4}$$

where n_{M_i} is the largest frequency in the i-the row, $max(C_j)$ is the largest column total and $\sum_{i=1}^{\prime m} n_{M_i}$ is the sum of all the maximum frequencies in the column associated with $max(C_j)$.

This is pretty useful because researchers have now the ability to perform inferential test. Because the statistic follow a normal distribution (Siegel & Castellan Jr, 1988), one may test if the lambda statistic is equal to some value using the familiar test:

$$z = \frac{\lambda_B - \lambda_{B0}}{\sqrt{var(\lambda_B)}} \tag{5.5}$$

	$STRE(t+1)$					
	1	2	3	4	5	6
$STRE(t)=1$	**12**	9	7	4	1	0
2	8	**45**	21	6	5	2
3	7	19	**42**	13	1	1
4	4	8	**10**	7	3	3
5	1	**5**	2	3	2	1
6	1	1	**2**	1	**2**	1
Total	33	87	84	34	14	8

Table 5.4: Transition frequency matrix of six-point scale stress (in bold typeface are the largest frequencies for each column)

Example Let's see how this lambda statistic applies. To illustrate it we use the transitional frequency matrix of the 6-scale stress variable. It is reproduced in table 5.4 [4].

The maximum frequency for each of the rows (state at time t) is respectively 12, 45, 42, 10, 5, 2 for a total of $\sum_{i=1}^{6} n_{M_i} = 116$. The column totals are 33, 87, 84, 34, 14 and 8; the maximum for these totals is thus in the second column, yielding $max(C_j) = 87$. And since the total number of transitions is 260, we are ready to compute the statistics [5]:

$$L_B = \frac{116 - 87}{260 - 87} = 0.167 \tag{5.6}$$

This result means that there is a reduction of 16.7% in the error when trying to predict the next stress state when knowing what state the person is in.

Computing now the variance for this stress variable we get (Siegel & Castellan Jr, 1988):

$$S(L_B) = \frac{(260 - 116)(116 + 87 - 2 * 45)}{(260 - 87)^3} = 0.0031 \tag{5.7}$$

The most evident test to be performed would be to determine if the reduction of uncertainty is really different from 0; however this can not be performed (Siegel & Castellan Jr, 1988). But one could test if the value is really different from say, lambda=10%. We thus have:

$$z = \frac{0.167 - 0.10}{\sqrt{0.0031}} = 1.20 \tag{5.8}$$

This value being inferior to the threshold of $z = 1.96$ at $\alpha = 0.05$, we can not reject the hypothesis that the decrease of error is greater than 10%.

[4]Produced with the S-Plus command:

```
>  TFM(STRE)
```

[5]Produced with the S-Plus command:

```
> Lambda(STRE)
```

Discussion Goodman and Kruskal's lambda statistic was presented in this section. It is an interesting measure of association, because it informs about the reduction of uncertainty when knowing what is the antecedent of the system. It ranges from 0 to 1, the former meaning no reduction of uncertainty and the latter full knowledge about the next state of the system.

Its main advantage relates to its asymmetrical nature. Contrary to other suggested tests, the way variables are paired is of utmost importance; rows and columns are not interchangeable. There is in fact a lambda statistic for the row-to-column association, and another for the column-to-row association. This is not the case of say, X^2-statistic, for which rows and column may be permuted and still giving the same results. Another advantage is the absence of constraints on the distribution of the variables; there are no rules stating that frequencies must be greater than 5, such as X^2 statistic. Moreover it can be performed on categorical variables, and consequently on configurations.

5.1.3 Chi-square and likelihood ratio statistics

Perhaps the state of the system at a certain time is completely independent from what happened previously. Or may be it does depend on the past state. How do investigators separate these two cases?

A well-known tool to investigate whether the present state of the system depends on the previous state is the chi-square statistic (Bishop, Fienberg, & Holland, 1975; Everitt, 1992). This statistic is usually employed on structural contingency tables in order to determine if there is a dependence between 2 (or many) categorical variables. As was shown by scientists, this statistic may well be applied on transitional frequency matrices, a special case of contingency tables (Bishop et al., 1975; Gottman & Kumar Roy, 1990; Bakeman & Gottman, 1986).

The chi-square statistic tests whether transitions are a *first-order* Markov process, that is, if the state of the system at time t depends on its state at time $t-1$, or if state at time $t+1$ depends on state at time t. *Higher order* Markov processes refer to a dependence at higher lags ($t-2$, $t-3$, and so on; they are reviewed in chapter 8).

Researchers can test four different types of hypothesis using chi-square statistics (Castellan, 1979):

1. fit of frequencies to specified probabilities
2. fit of marginal frequencies to specified probabilities
3. conditional fit of row (or column) frequencies to specified probabilities
4. variables independence

We will discuss the fourth type of hypothesis and let readers refer to Castellan for more details about the three others, since they are less frequent.

Method For a sequence of m modalities, there are $m \times m = m^2$ first-order transitional frequencies, as represented by the general transitional frequency matrix (cf table 3.1). These observed transitional frequencies are compared with what would be expected if the state of the system at time $t+1$ was independent from its previous state. The best estimates are given by the maximum likelihood estimates (Bishop et al., 1975; Everitt, 1992):

$$E_{ij} = N\hat{p}_{i+}\hat{p}_{+j} = N\frac{n_{i+}}{N}\frac{n_{+j}}{N} \tag{5.9}$$

$$= \frac{n_{i+}n_{+j}}{N} \tag{5.10}$$

These expected frequencies are then replaced in the well known X^2 formula:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \tag{5.11}$$

A less known statistic, the likelihood ratio G^2, may also be used for testing independence between categorical variables:

$$G^2 = 2\sum_{ij} n_{ij} \log \frac{n_{ij}}{E_{ij}} \tag{5.12}$$

with $(m-1)^2$ degrees of freedom.

The two statistics X^2 and G^2 have asymptotically the same behavior. Should they differ too greatly, such as with sparse tables, researcher is advised to be cautious about the validity of the results.

Hypothesis When using the X^2 test, investigators make the following hypothesis about the transitional frequency matrix:

$$H_0 : \text{x(t) is independent from x(t-1), e.g, } n_{ij} = \frac{n_{i+}n_{+j}}{n_{++}} \tag{5.13}$$

$$H_1 : \text{x(t) is dependent from x(t-1), e.g, } n_{ij} \neq \frac{n_{i+}n_{+j}}{n_{++}} \tag{5.14}$$

Researchers may test another hypothesis using the above formula. They may hypothesize that the conditional probabilities are the *homogeneous* (the same) between the rows (or columns). Whatever state i at time t is considered, the conditional probabilities for each state j at time $t+1$ are equal. That is: $p_{j|1} = p_{j|2} = p_{j|m}$. This is an important test because if there is a particular structure in the temporal dependence, there should be inequalities in these conditional probabilities.

The later test is in fact equivalent to the former because in both cases the expected values are equal. The expected frequencies when conditional probabilities are equivalent are given by:

$$E_{i,j} = n_{i+}\hat{p}_{j|i} = \frac{n_{i+}n_{+j}}{N} \tag{5.15}$$

The equation is the same as for testing the overall dependence, and it is the same for testing the conditional probabilities of rows or columns.

Interpretation If the hypothesis is rejected it is concluded that there is a structure of the data sequence, because there is a dependence between the state at time t and the state at time $t+1$. Otherwise the independence of states must be accepted.

The X^2 and G^2 tests may also be interpreted as tests of homogeneity. That is, the transitional probabilities are the same for all levels i of the variable.

Example For our example, using the dichotomized (BSTRE), trichotomized (STRE3) and full-scale (STRE6) stress variables, we test the independence of transitions using the X^2 and G^2 statistics. The result are in table 5.5.

Variable	X^2	G^2	df	p-value
BSTRE	13.82	15.29	1	< 0.005
BEMOT	25.74	28.54	1	< 0.005
BFAM	0.22	0.20	1	0.64
Config	82.09	60.94	36	< 0.005

Table 5.5: Tests of independence on transitions

The p-values are smaller than 0.05 for variables BSTRE, BEMOT and the configuration; We may therefore conclude that there is a dependence between the state of these variables at time t and at time $t+1$. The test on BFAM variable is not significant so we can not interpret the transitions as dependent. Using X^2 or G^2 statistics yield similar results.

The second interpretation of the test, in term of homogeneity, tells us that the conditional probabilities of state BSTRE=0 (0.84 and 0.16) are not equal to the conditional probabilities of state BSTRE=1 (0.60 and 0.40); the same reasoning applies for the other variables. However we must be cautious with the configuration variable because there were many transitions frequented less than 5; there are 20 transitions (out of 36) that appeared less than 5, so the test is not reliable enough in this case.

Discussion The use of the X^2 or G^2 tests in case of transitional frequency matrices follows the same principles as with static contingency tables (Everitt, 1992). One has to beware of transition cells less than 5.

The statistic must be used with caution. A significant p-value does not indicate at all the *type* or *direction* of dependence between the states. The investigator should always examine the transitional frequency matrix before, and relate it to the significance of the test. Two p-values may be significant, but one may refer to a tendency to reduce the variable levels, the other to maintain it high; although both significant, these are very different processes.

Another caution that applies to this statistic as well as other ones, is that it concerns only the *first-order* dynamics. A non-significant statistic may show that there is no dependent $t \rightarrow t+1$ transitions, but that the dynamics is a second-order dynamics. One has to investigate further (cf chapter 8.5).

Conclusion There is finally a test that allows to determine if the sequence of observations depends on the past state of the system. Investigators do not have to rely only of counts and "simple" probabilities. The X^2 or G^2 tests are appropriate statistics, if one follows its principles and prerequisites. Their limitations were discussed in this section.

These tests apply to the overall dynamics of the system. Investigators willing to examine a specific transition should consult the following section (5.2).

5.2 Testing specific transitions

5.2.1 The test of proportions

Instead of testing the whole transitional frequency matrix as in the last section, investigators may be interested in specific transitions that occurred during the period of observation. We present here the essential tests available for answering these questions.

What we want to determine is when the system is in some state $X(t) = i$, is the transition to another state $X(t+1) = j$ significant? Indeed it is insufficient to just compute the transitional probability and see if it is greater than the equiprobable probability. For example, in a dichotomized variable, one would expect to get a probability of 0.50 if transitions were equiprobable. But how to tell if a probability of say, 0.52 or 0.61, is really greater than 0.50? A sound statistical test is required.

Method Statistical literature suggests a number of possibilities regarding how to compute the significance of a transition. Let's begin with the case of a 2×2 transitional frequency matrix. One of the simplest choice is to test, for a given state, whether the transitions toward each two states are equiprobable; that is: $p(i \to i) = p(i \to j) = 0.50$. So given a state i, there are then equal chances to transit to state i or state j. It then implies there is no significant dynamics underlying the sequence of states. The appropriate statistical test for this operation is the well-known binomial test or test of proportion (Spiegel, 1980), as in equation 5.16:

$$z = \frac{P - p}{\sqrt{pq/n}} \tag{5.16}$$

where P is the equiprobability value to be tested, p the observed probability, $q = 1 - p$ and n the number of transitions. The computed z value is to be compared with a z table for assessing its significance.

Because testing of multiple transitions is often involved here, researchers should follow Bonferonni's rule to reduce the chance of rejecting hypotheses: the usual $\alpha = 0.05$ significance level is divided by the number of tests performed.

Hypothesis The following hypothesis is tested for the transition p_{ij}, where the variable has m levels:

$$H_0 : p(i \to j) = \frac{1}{m} \tag{5.17}$$

$$H_1 : p(i \to j) \neq \frac{1}{m} \tag{5.18}$$

	$BSTRE(t+1)$	
	0	1
BSTRE(t) = 0	0.84	0.16
1	0.60	0.40

Table 5.6: Matrix of probability of transitions for dichotomized stress variable

Transition	X^2	d.f.	p-value
BSTRE 0 $\rightarrow$ 0	91.11	1	< 0.005
BSTRE 1 $\rightarrow$ 0	1.75	1	0.19

Table 5.7: Test of proportion for dichotomized stress variable

Interpretation The larger the departure from theequiprobability the better the chance transitions are significant.

Example Now let's apply this test to our stress example, considering the dichotomized case first. Table 5.6 presents the transitional frequency matrix [6] and table 5.7 [7] [8] shows the results of statistical tests.

Results show that there is a significant departure from theequiprobability for transitions from state BSTRE=0 (p-value < 0.005), but not for BSTRE=1 (p-value = 0.19). Therefore we can say here that when stress is low, the probability of transiting towards a low stress is meaningfully different (greater) than to a high stress; for a high level of stress, although the probability towards a low stress is greater than for a high level, it is not significantly higher.

Let's now consider the dynamics of a categorical variable, namely PLACE, for another subject, XR3. Table 5.8 shows her transitional frequencies. The table shows clearly that she has a tendency to remain in the same state the next moment, especially when she is at home (there are 64 such transitions), but not when she is in a public place (thus implying that this subject almost never stays for more than 4 consecutive hours in a public place). A question of pragmatical importance is what typically happens next when she is at work. The most frequent transition is towards work itself ($n = 13$), but what about towards home or an other place? Are these transitions more likely than the uniform probability? Let's investigate.

We now test the 3 transitions from work to work, from work to home and from work

[6] Produced with the S-Plus command:

```
> TPM(BSTRE)
```

[7] Produced with the S-Plus command:

```
> EquiprobabilityTest(TFM(BSTRE), fromstate="0", tostate="0")
```

[8] Produced with the S-Plus command:

```
> EquiprobabilityTest(TFM(BSTRE), fromstate="1", tostate="0")
```

	$PLACE(t+1)$				
	Friends	Home	Other	Public	Work
$PLACE(t)$ =Friends	**27**	4	5	4	0
Home	4	**64**	21	3	9
Other	5	20	**35**	5	6
Public	**5**	4	4	1	2
Work	0	8	6	3	**13**

Table 5.8: Transitional frequency matrix for PLACE variable (subject : XR3; in bold the highest transitional frequency for each state i)

Transition	X^2	d.f.	p-value
Work → Home	0.47	1	0.49
Work → Other	0.00	1	1.00
Work → Work	8.80	1	0.003

Table 5.9: Test of proportion for PLACE variable (subject: XR3)

to an unspecified place ("other"); statistical results are in table 5.9 [9] [10] [11]. Because multiple testing is also involved here, following Bonferonni's rule the usual $\alpha = 0.05$ significance level is divided by 3 (to $\alpha = 0.017$) to reduce the chance of errors. Here the theequiprobability against which the test is performed is $P = 0.2$, since there are 5 different levels.

Results show that there are only one transition that is significantly different from the equiprobable value of $P = 0.2$, the transition from work to work (its significance level $p = 0.003$ is lower than $\alpha = 0.017$). The other transitions should not be considered as different from the theequiprobability. We may then conclude that this subject has a greater tendency to keep on working when she is already at work.

Discussion A first procedure for testing specific transitions was presented in this section. It is the test of proportion, where one tests a given conditional probability with respect to theequiprobability.

Its advantage is that it is easy to understand and compute. Its main limitation comes from the fact that it does not take into account the marginal frequency of the given state; states that occurred 5 times or 1200 times are treated similarly.

Investigator should pay attention to the interpretation of the result. A significant result

[9] Produced with the S-Plus command:

```
> EquiprobabilityTest(TFM(PLACE), fromstate="Work", tostate="Home")
```

[10] Produced with the S-Plus command:

```
> EquiprobabilityTest(TFM(PLACE), fromstate="Work", tostate="Other")
```

[11] Produced with the S-Plus command:

```
> EquiprobabilityTest(TFM(PLACE), fromstate="Work", tostate="Work")
```

does not indicate the *direction* of the significance. The tested probability may be significantly more – or less– than theequiprobability. One has to examine the transitional frequency matrix to know it.

Conclusion Testing theequiprobability is one step towards the realization of our objective. But a better test should take into account the unconditional probability of the states for computing the significance of the transition. The next sections explain these kinds of procedures.

5.2.2 Sackett's z-score for testing transitional probabilities

Sackett, followed by Gottman, suggested a z-score, by comparing the conditional probability of a transition to its unconditional probability. It informs on the reduction of uncertainty given the knowledge of the past state on the present one. Contrary to the test of proportion, this method takes into account the number of transitions observed into a sequence.

Method Sackett's formula is written as in equation 5.19:

$$z_s = \frac{p_{j|i} - p_j}{\sqrt{\frac{p_j(1-p_j)}{(n-k)p_i}}} \tag{5.19}$$

where n is the total number of transitions and k, the lag to be analyzed. It has an asymptotically normal distribution for large n ($n > 20$) and variance of one.

In the numerator is the difference of proportion and the denominator is the estimated standard error. But, as Allison and Likert pointed out (Allison and Likert, 1982, p. 394), the denominator is not the correct standard error of the numerator. Allison and Likert argued it would be correct if p_j were the true probability, and not merely an observed proportion subject to sampling error. They also remark that Sackett's formula (5.19) is too conservative. The corrected (and correct) formula they suggest is rather:

$$z_s = \frac{p_{j|i} - p_j}{\sqrt{\frac{p_j(1-p_j)(1-p_i)}{(n-k)p_i}}} \tag{5.20}$$

It relates to Sackett's formula by way of:

$$z_1 = \frac{z_s}{\sqrt{1 - p_i}} \tag{5.21}$$

The Allison and Likert formula was shown to be equivalent to the square root of a chi-square test (Allison & Liker, 1982; Fienberg, 1980). When applied to a 2×2 matrix, it is equivalent to a test of proportion.

Remember to use Bonferonni's rule if more than one specific transition is tested.

Example Let's apply the formula on the transitions of our dichotomous stress example. We want to test if the reduction of uncertainty brought by conditional probabilities on low and high stress ($p_{1|0}$ and $p_{0|1}$) are significant. Replacing the empirical values in equation 5.20, we get:

Transition	Sackett's z	p-value
Work → Home	-1.46	0.14
Work → Other	-0.95	0.34
Work → Work	5.76	< 0.005

Table 5.10: Test of proportion for PLACE variable (subject: XR3)

$$z_{1|0} = \frac{0.16 - 0.22}{\sqrt{\frac{0.22(1-0.22)(1-0.78)}{(261-1)0.78}}} = -4.12 \tag{5.22}$$

$$z_{0|1} = \frac{0.84 - 0.78}{\sqrt{\frac{0.78(1-0.78)(1-0.22)}{(261-1)0.22}}} = -3.82 \tag{5.23}$$

The two values are well above the threshold of 1.96, for an α of 0.05. The results imply that the knowledge of the past state bring a significant increase of information about the present state.

The result of this test on binary variables is symmetrical, but with a reversed sign. That is, if we test the transitions $p_{0|0}$ we would get $z = 4.12$ and for $p_{1|1}$ we would get $z = 3.82$

Looking now at the dynamics of PLACE variables, subject XR3, as in table 5.8. We will test once again the transitions from work to work, from work to home and from work to an unspecified place ("other"); statistical results are in table 5.10 [12] [13] [14].

Here results are similar to the tests on equiprobable transitions (but this is not always the case, of course). It thus indicates that the conditional probability from work to work is significantly different from the unconditional probability; the other two transitions are not significant.

Discussion The z-score test presented here is a good candidate for the statistical analysis of specific transition.

Results are not always the same as the tests of equiprobable transition, because they do not test the same thing. The equiprobable test examines if a conditional probability is statistically different from uniform probability, where the other one tests the difference between the conditional probability with the unconditional probability.

Conclusion The z-score test received both confirmations and criticisms, but corrected formula were suggested so provide better estimates of the asymptotic standard error (Allison

[12] Produced with the S-Plus command:

```
>  SackettZ(PLACE,fromstate="Work",tostate="Home")
```

[13] Produced with the S-Plus command:

```
>  SackettZ(PLACE,fromstate="Work",tostate="Other")
```

[14] Produced with the S-Plus command:

```
>  SackettZ(PLACE,fromstate="Work",tostate="Work")
```

& Liker, 1982). Consequently it seems this z-score test is an appropriate test for specific transition dependences. We advise users of this test to always draw a state transition diagram to visually summarize significant transitions. Also since a p-value does not inform about the type (direction) of dependence, researchers should always examine the numerical values of both conditional and unconditional probabilities.

5.2.3 Chi-square based tests

In the last sections two major tests were presented for examining the significance of specific transitions. Researchers proposed other ways to test this significance. We review some of them here.

Method Some techniques directly relate to the X^2-statistic. No matter if a X^2 test on the whole transitional frequency matrix is significant or not, researchers may examine specific transitions. It is indeed often a good approach to examine the fit of individual cells in contingency tables (Bishop et al., 1975; Eye, 1990). This allows to detect the source of the overall significance.

One main technique is to compute a z-score based standardized cell residuals, by comparing the observed and expected value of the cells. Three summary measures are proposed (Bishop et al., 1975):

1. the component of X^2, given by
$$z_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}} \tag{5.24}$$
2. the component of G^2, given by
$$z_{ij} = 2n_{ij} \log \frac{n_{ij}}{e_{ij}} \tag{5.25}$$
3. the Freeman-Tukey deviates:
$$z_{ij} = \sqrt{n_{ij}} + \sqrt{n_{ij}+1} - \sqrt{4e_{ij}+1} \tag{5.26}$$

They all behave like a normal distribution of mean 0 and variance 1. To determine the significance, investigators compare the value of computed z-score in a table of z, at a given α level.

Example We will show here the X^2 test of specific transitions; the G^2 test is left aside here, because of the great similarity with X^2 statistics. The tests are performed on each of the following variables: BSTRE, PLACE and configuration composed of BFAM, BEMOT and BSTRE.

Let's start with the dichotomized stress variable, BSTRE. The X^2 residuals computed on its transitional frequency matrix are given in table 5.11 [15]. Since these X^2 residuals

[15] Produced with the S-Plus command:

```
> Residuals(BSTRE)
```

may be interpreted as z-scores, a transition is considered significant when its value exceeds $z_{\alpha=0.05} = 1.96$. Because of the multiple testing, the α-level should be decreased accordingly; four transitions are tested here, so $\alpha = 0.05/4 = 0.0125$, for a $z_{\alpha=0.0125} = 2.50$.

	$BSTRE(t+1)$	
	0	1
$BSTRE(t)$ = 0	0.85	-1.60
1	-1.60	**3.06**

Table 5.11: X^2 residuals on transitional frequency matrix, BSTRE (in bold typeface is the significant transition)

Only one transition has its residual score exceeding the $z = 2.50$ value and must be considered as significant: when the subject remained in a high level of stress $BSTRE_{1\to1}$. All others transitions are non-significant. It is somewhat contrary to what we expected: the transition $BSTRE_{0\to0}$ seemed much more important than the other ones. But let's remind the reader what is tested: the difference between what is actually observed and what is expected under the hypothesis of independence between the antecedent state and its consequent. Transitions $BSTRE_{0\to0}$ is more frequent than the others but it does not significantly depart from the independence model.

Let's examine now specific transitions from the PLACE variables (subject XR3, table 5.12 [16]). The multiple testing of transitions requires to lower the α-level at $0.05/25 = 0.002$, which corresponds to a z-score of 3.09. Any residual greater than 3.09 implies a transition significantly different from the expectation under a model of independence, and conversely.

	$PLACE(t+1)$				
	Friends	Home	Other	Public	Work
$PLACE(t)$ = Friends	**8.19**	-2.92	-1.81	0.96	-2.16
Home	-3.01	**3.97**	-1.29	-1.30	-0.80
Other	-1.87	-1.43	**3.50**	0.28	-0.79
Public	1.54	-0.88	-0.19	0.01	0.10
Work	-2.18	-1.06	-0.79	0.84	**5.09**

Table 5.12: X^2 residuals on transitional frequency matrix, PLACE (subject: XR3; in bold are significant transitions)

Interestingly there are four transitions that significantly stray from expectations, and they all refer to transitions towards themselves: Friends, Home, Other and Work. It implies that the subject significantly stayed in the same place for two consecutive periods of observations.

[16]Produced with the S-Plus command:

```
> Residuals(PLACE)
```

Stability in the dynamics of place of interaction is therefore a salient characteristic for this subject.

Discussion We presented three procedures for testing specific transitions on a transitional frequency matrix. They are based on expected values as computed with usual contingency table techniques. The tests are the analysis of residuals of X^2 and G^2 statistics, and the Freeman-Tukey deviates. So investigators interested in testing specific transitions may well consider only one of these tests, the X^2 analysis of residuals probably being the more appropriate choice, since it is better known.

Researchers should pay attention to the α level of significance. The multiple testing procedures should implied the use of a lower α level, so to reduce the risk of rejecting hypotheses just by chance (Eye, 1990). The Bonferonni criteria provides a good reference, where the α level is divided by the number of tests to be performed. For example, a 6x6 transitional matrix implies the computation of 36 individual tests; therefore one would set the α level at $0.05/36 = 0.0014$

5.3 Conclusion

Investigators have in their hands a powerful arsenal for testing both the overall transitional dependences of the transitional frequency matrix and specific transitions. Conjugated with graphical representations such as dynamical Karnaugh map and state transition diagrams, the analysis of dynamics is now both scientifically sound and easily communicable with only the most interesting transitions being presented.

The following chapters will help determine the order and complexity of the dynamics, by mainly referring to information theory (chapter 6.1). Phases, higher order Markov chains and modelling of dynamics will be analyzed afterwards.

Chapter 6

Entropy and complexity

Summary. *Information theory, once a fashionable trend in psychology during the years '50–'70 but forgotten for a while, has recently received a renewal of attention. Latest developments in complex dynamical systems constantly refer to this fruitful framework for describing and quantifying the dynamics of systems. We use it for assessing the amount of order and disorder in psychological systems, complexity of transitions, as well as determining the past states influences on systems.*

Information theory was at first considered a subset of communication theory (Shannon & Weaver, 1949; Cover & Thomas, 1991). Its main purpose was to answer two fundamental questions: 1) what is the ultimate data compression that can be applied to a signal and 2) what is the ultimate transmission rate of signals on a wire? But the mathematical techniques developed so fruitfully that they were applied in various fields of investigation: in statistical physics (thermodynamics), computer science (Kolmogorov and algorithmic complexity) and probability and statistics (hypothesis testing), to name but a few.

6.1 Information theory: a primer

The definition of information theory is controversial. Who would settle for a definition such as "the amount of uncertainty contained in a symbol"? (Shannon & Weaver, 1949). What about the information of a document? of a web page? "It is very difficult to make a general theory of information, and very easy to give a bad answer to this question. Shannon and Weaver's developments have pretended to have the key to the problem and the intimidation have worked so well that many believe that information is what this theory is about and nothing else" (Delahaye, 1994). Nonetheless researchers have accepted this focused definition and developed a set of mathematical tools for analyzing the information of systems (McCulloch & Pitts, 1974).

Suppose a person asks a question that can be answered by yes or no. There is an uncertainty about the possible answer; if it were completely certain, it would not be necessary to ask the question… A yes/no question (with uncertain outcome) carries 1 bit of information. In general, a question with N possible answers carries $\log_2 N$ bits of information. So the higher the number of possibilities, the higher the uncertainty, thus reflected in a higher degree of information.

If there is knowledge about the possible outcomes, then the amount of information is reduced accordingly. A question with 8 possible answers for which the questioner knows that 6 of them are impossible reduces his uncertainty about the answer (leaving two uncertain outcomes). Therefore one simply subtract the information before and after, yielding $\log_2 8 - \log_2 2 = 2$ bits of information.

Extending the previous reasoning we see that having knowledge regarding a question (or about a the state of a system) reduced the risk of making a wrong prediction.

$$\log_2 N_{before} - \log_2 N_{after} = \log_2 \frac{N_{before}}{N_{after}} = -\log_2 \frac{N_{after}}{N_{before}} = -\log_2 P \tag{6.1}$$

where P is the probability of guessing right. Suppose one knows 4 answers out of 8 for a question, the information is then $-\log_2 4/8 = 1$ bit of information left.

The use of logarithms instead of direct numbers makes bits *additive* when options are *multiplicative* (recalling that $\log_b XY = \log_b X + \log_b Y$). Asking two independent questions should provide an independent amount of information. This is true since

$$-\log_2 P_{XY} = -\log_2 P_X P_Y \qquad = -\log_2 P_X \log_2 P_Y = I(X) + I(Y) \tag{6.2}$$

6.1.1 Entropy

Shannon (Shannon & Weaver, 1949) derived a formula for the amount of information contained in a signal. It is named entropy, after the previous works of Boltzmann in thermodynamics. His approach consisted in setting a number of desirable properties that a function should have in order to quantify the amount of information. He reasoned in terms of uncertainty or surprise that a given output of a system brings to an observer. The information should be zero when an observer knows for sure the output of a system, and be maximal when any output is equally likely. Shannon derived the formula of entropy.

Method Let x be a discrete variable which may take $i = 1...m$ possible values. Entropy is computed as follows (equation 6.3):

$$H(x) = -\sum_{x=i}^{m} p_i(x) \log_2 p_i(x) \tag{6.3}$$

The entropy varies from 0 to $\log(m)$. When the system takes only 1 value or symbol, then the entropy is $H(x) = 0$ (by definition $\log 0 = 0$). When the system may take all m possible values with equal probability then then entropy is maximal, at $H(x) = \log(m)$. It is obvious that the greater the number of possible states the greater the maximum entropy. If a system may take $m = 2, 4$ or 16 possible values, the maximal entropy is $H = 1, 2$ or 4 respectively.

Entropy can be seen as the average amount of information required to select observations by categories (Krippendorff, 1986).

Entropy may be *standardized* so that it would range from 0 to 1, by dividing it by its maximum $\log(m)$. It may be easier to compare the amount of disorder of two systems, knowing that one system encountered more states than the other.

The base of the logarithm often taken is 2. The resulting measure is then said to be in *bits* . One may take another base for the computation, for example a base e, and the measure is said to be in *nats* , or a base 10. One may transform an entropy computed in a certain base a to another base b by multiplying the given entropy by $\log_b a$. The choice of the base is relatively arbitrary, but often taken according to the data scale.[1]

A nice property of entropy is that variable categories may be permuted without changing its value. Only the relative frequencies matter. This is why this measure is said to be *content-free* . It does not make any assumptions about the distribution of data; it thus belongs to the *nonparametric* family of statistics.

Interpretation Entropy is interpreted in many different ways. It is a measure of the *uncertainty* tied to the observed system. The lower the entropy the easier it is to predict the system's state, and conversely. It may also be interpreted as a measure of *disorder* of a system, or in a very similar fashion, a measure of its *variability* . Again, the lower the entropy the more orderly the system (and the less variable), and conversely, the higher the entropy the more disordered and more variable the system (because it may take all possible values with equal probability).

Example Here is the distribution of a variable (here the stress distribution of our subject FA7, table 6.1), along subsequent steps of the entropy computations:

Category	Freq	Prob	$\log_2(Prob)$	Prob * $\log_2(Prob)$
1	33	0.13	-2.98	-0.38
2	87	0.33	-1.58	-0.53
3	84	0.32	-1.64	-0.53
4	35	0.13	-2.90	-0.39
5	14	0.05	-4.22	-0.23
6	8	0.03	-5.03	-0.15
Sum	261	1.00	-18.35	-2.20

Table 6.1: Frequency table for the STRESS variable

The entropy of this variable is computed by multiplying the probability of each category by its logarithm, summing them over and taking the negative value of this sum. Here the entropy is $H(STRE) = 2.20$.

There is an uncertainty of 2.20 bits when someone observes the stress of this subject. It means that it would require on average a little more than two questions to guess in what state the subject is.

Figure 6.1 presents the entropy of 7 bio-psycho-social full-scale variables of our subject example. It mainly shows that the entropy of these variables range from 1.59 (FAM vari-

[1]There should be a constant K in the equation that reflects the arbitrary choice of base, but is usually left off.

able) to 2.20 (STRE) variable. It thus appear that the familiarity of situations produced more concentrated answers than the stress variable.

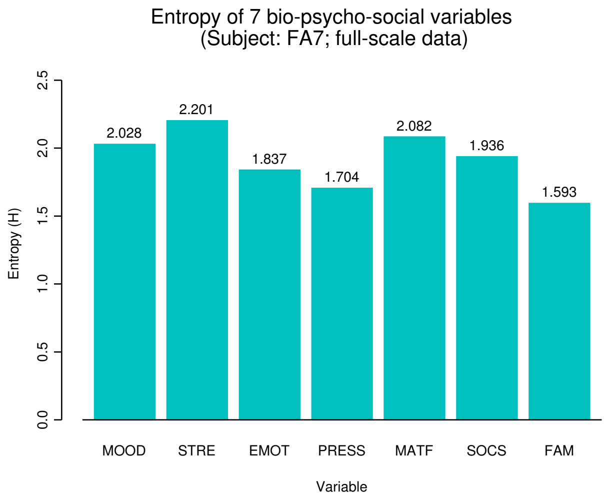

Figure 6.1: Entropy of 7 bio-psycho-social variables

Indeed when the distribution of these variables are examined (figure 6.2), both variables have two levels that are well-represented (STRE=2 and 3, FAM=4 and 5); familiarity of situations was however evaluated more frequently by the same levels that were the stress variable, which explains the lower entropy of the former variable.

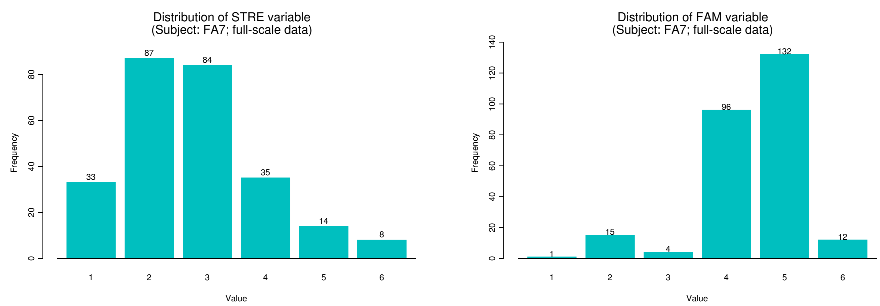

Figure 6.2: Distribution of STRE and FAM variables

Recoding variables There are times when recoding variables becomes necessary, as was previously shown in chapter 2 for building configurations. For example we can easily recode the 6-point scale stress variable into a 3-point scale and a 2-point scale variable.

Let's do the trichotomisation of the variable (named STRE3) following this scheme:

Full scale	Trichotomized scale
$1 \leq STRE \leq 2$	1
$3 \leq STRE \leq 4$	2
$5 \leq STRE \leq 6$	3

and its dichotomisation into BSTRE was performed following this scheme:

Full scale	Dichotomized scale
$1 \leq STRE \leq 3$	0
$4 \leq STRE \leq 6$	1

The distribution of these variables is given in figure 6.3. Levels 1 and 2 of the trichotomous stress were encountered 120 and 119 times (for both a corresponding probability of 0.46), while level 3 was encountered 22 times (for a probability of 0.08). Regarding the dichotomous stress variable, the low stress level BSTRE=0 was seen 204 times, accounting for 78% of all states, leaving 22% for the higher stress level.

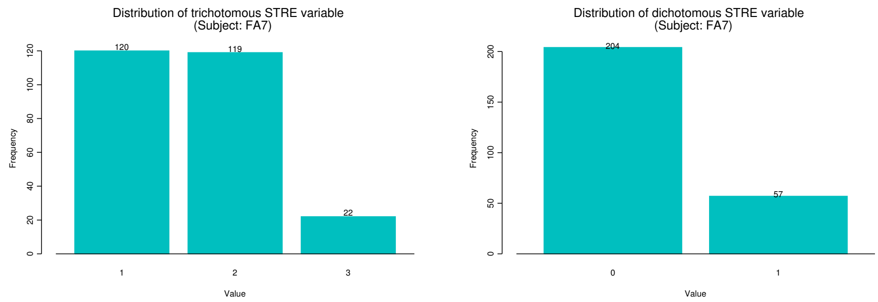

Figure 6.3: Distribution of trichotomous and dichotomous STRESS variables

From these frequencies and probabilities one easily computes the entropy of the new variables. $H(STRE3) = 1.33$ bits and $H(BSTRE) = 0.76$ bits; table 6.2 summarizes the results. For comparing entropies of variables of different scales it is somewhat preferable to measure the amount of *standardized entropy* . It is simply calculated by dividing the entropy with its maximum entropy, $\log_2(m)$. The maximum entropy for a 2, 3 and 6 scale variable is $\log_2(2) = 1$, $\log_2(3) = 1.58$ and $\log_2(6) = 2.58$ respectively. It is no coincidence that $\log_2(6) = \log_2(2) + \log_2(3)$, since $\log(a * b) = \log(a) + \log(b)$. Standardized entropy ranges from 0 to 1 (or 0% to 100%).

Consequently the standardized entropy for the stress variable using a 2-, 3- and 6-point scale is 0.76 (0.76/1), 0.84 (1.33/1.58) and 0.85 (2.20/2.58). It thus implies that although the 3-point scale employs half the number of categories of the original scale, the amount of disorder (or uncertainty) is almost identical. This suggests that there are categories in

Variable	Entropy	Standardized Entropy
STRE (full scale)	2.20	0.85
STRE3 (trichotomous)	1.33	0.84
BSTRE (dichotomous)	0.76	0.76

Table 6.2: Entropy for three STRESS variables

the 6-point scale that are not much frequent, and that recoding into a 3-point scale would not drastically reduce the amount of information. This avenue is numerically explored in section 6.2.

Discussion The entropy quantity has been known for a long time now. In statistics it is the *de facto* measure of the dispersion of a categorical variable. Unfortunately it is rarely taught or described in statistical textbooks; for example, see its absence in Howell (1997) and Saporta (1990). It seems once again that categorical variables are thought as less than important and put aside. We hope this work will bring an increased interest in this simple, adequate yet powerful measure.

6.1.2 Joint entropy

Shannon's entropy works very fine for describing the order, uncertainty or variability of a unique variable. But what about entropy for more than one variable? There are various entropies when considering 2 or more variables together: the joint entropy, the mutual information and the conditional entropy. We first review the joint entropy.

Method When considering two discrete variables x and y at the same time, it is possible to measure the degree of uncertainty or information associated with them. It is called the *joint entropy* , $H(x, y)$. If x and y may respectively take m_1 and m_2 possible values, the joint entropy is computed as in equation 6.4:

$$H(x,y) = -\sum_{x=i}^{m_1}\sum_{y=j}^{m_2} p_{ij}(x,y)\log_2 p_{ij}(x,y) \tag{6.4}$$

where p_{ij} represents the probability of being classified in both category i of variable x and category j of variable y.

The joint entropy varies from a theoretical 0 (or empirically $min(H(x), H(y))$) to $\log_2(m_1) + \log_2(m_2)$. The relation between the individual entropies and their joint entropy is given by equation 6.5:

$$H(x,y) \leq H(x) + H(y) \tag{6.5}$$

It expresses the fact that the joint entropy is always smaller then the sum of the individual entropies. The equality holds only when the two variables are independent.

Example Let's examine the joint entropy of two variables, the stress (STRE) and emotionality (EMOT) of our subject example. We present first their contingency table and their joint probabilities (tables 6.3 and 6.4).

	STRE					
	1	2	3	4	5	6
EMOT= 1	19	4				
2	11	63	64	3	1	
3	2	16	18	20	2	2
4	1	4	1	9	6	2
5			1	2	4	3
6				1	1	1

Table 6.3: Contingency table for stress and emotionality

	STRE					
	1	2	3	4	5	6
EMOT= 1	0.07	0.02				
2	0.04	0.24	0.25	0.01		
3	0.01	0.06	0.07	0.08	0.01	0.01
4		0.02		0.03	0.02	0.01
5				0.01	0.02	0.01
6						

Table 6.4: Joint probabilities for stress and emotionality

In the latter table we multiply each probabilities with its logarithm, sum them over and take the negative of this number. The result is the joint entropy: $H(EMOT, STRE) = 3.46$ [2]. It means that on average it would requires more than 3 (and less than 4) questions to guess the level of both variables.

We further compute the entropies of our binary variables, adding a third variable, BFAM, the perceived familiarity of the situation. Table 6.5 show how these variables combined together.

It shows that the combination of variables that produce the smallest amount of entropy is BFAM - BEMOT ($H = 0.93$), while the highest joint entropy is for BSTRE - BEMOT ($H = 1.14$). The combination of the three variables yields an entropy of 1.46. Unfortunately these quantities does not lead to a direct interpretation because they do not have the same boundaries. The minimum entropy of the joint variables is given by the minimum of the two entropies and, the maximum is given by multiplying the log of the number of categories

[2] Produced with the S-Plus command:

```
> JointEntropy(EMOT,STRE)
```

Variables	Entropy (H)
BSTRE - BEMOT	1.14
BSTRE - BFAM	1.09
BFAM - BEMOT	0.93
BFAM - BEMOT - BSTRE	1.46

Table 6.5: Joint entropy for three binary variables

for each variable. We will see that the next measure, mutual information is more readily interpretable.

6.1.3 Mutual information

The next important measure based on entropy is *mutual information* . It is a measure of the information *shared* by variables, or the quantity of information an observer gets common in two (or more) variables.

Method There are many ways of expressing its formula. It may be expressed as a relation between the individual entropies and the joint entropy. It is the sum of the individual entropies, minus the joint entropy, as expressed in equation 6.6:

$$I(x, y) = H(x) + H(y) - H(x, y) \tag{6.6}$$

For the three variable case, the equation becomes:

$$I(x, y, z) = H(x) + H(y) + H(z) - H(x, y) - H(x, z) - H(y, z) + H(x, y, z) \tag{6.7}$$

We have previously seen that if the two variables are independent, the sum of their individual entropies is equal to the joint entropy. Then the mutual information is 0.

Example For our example, we compute the mutual information by replacing the x and y variable by EMOT and STRE, to yield:

$$I(EMOT, STRE) = H(EMOT) + H(STRE) - H(EMOT, STRE)$$
$$I(EMOT, STRE) = 1.84 + 2.20 - 3.46 = 0.58 \text{ bits}$$

Those two variables seem not to have a lot of information in common, only 0.58 bits of information. This number can be re-expressed in terms of scale, using $2^{0.58}$, which is 1.49. That means they have 1.49 unit scales in common, a lot more than previously thought.[3].

We then proceed to compute the mutual information for our group of three binary variables (cf. table 6.6 [4] [5]).

[3]The Pearson correlation between those two variable is $r = 0.68$

[4]Produced with the S-Plus command:

Variables	Mutual Information (I)
BSTRE - BEMOT	0.19
BSTRE - BFAM	0.06
BFAM - BEMOT	0.04
BFAM - BEMOT - BSTRE	0.19

Table 6.6: Mutual information for three binary variables

The last table exhibits the shared information between pairs and triplets of variables. The pair sharing the most information is BSTRE - BEMOT, while the least is BFAM - BEMOT. When combining the 3 variables together the mutual information amounts to the same as with the BSTRE-BEMOT configuration; therefore in this case the BFAM variable does not share much information with the 2 others.

6.1.4 Conditional entropy

Not only can we measure information about two variables as a whole, but also the amount of information of a variable *knowing* the other. This is called the *conditional entropy* . The higher the conditional entropy the more an observator can predict the state of a variable, knowing the state of the other variable. While we present in this section the computation on a synchronical level, we delve into the dynamical level in section 6.3.1.

Method Conditional entropy relies on *conditional probabilities* , or as were previously named on transitional probabilities. Suppose we want to compute the conditional probability of state j of variable Y *given* state i of variable X; this is written as $p(Y = j|X = i)$. It is computed by dividing the frequency of occurrence of the two states n_{ij} by the total frequency of state i, n_i. The conditional probability is then $p_{j|i} = \frac{n_{ij}}{n_i}$. It is different from joint probability, which is calculated by dividing a frequency n_{ij} by the grand total N.

Example For example, let's compute the conditional probability of each category of stress *given* the categories of emotionality (table 6.7). Using the frequency table 6.3, each cell entry is divided by the row total.

We observe that when emotionality is very low (EMOT=1), the probability that stress is very low (STRE=1) is 83%, that it is low (STRE=2) 17%, and zero for all other categories. Knowing the emotionality is very low greatly reduces the uncertainty regarding the state of stress. Such analysis should be continued for all categories. But let's return to our matter, the conditional entropy.

The equation for computing the conditional entropy is (equation 6.8):

```
> MutualInfo(BSTRE,BEMOT)
```

[5]Produced with the S-Plus command:

```
> MutualInfo3(BFAM,BEMOT,BSTRE)
```

	STRE					
	1	2	3	4	5	6
EMOT = 1	0.83	0.17				
2	0.08	0.44	0.45	0.02	0.01	
3	0.03	0.27	0.3	0.33	0.03	0.03
4	0.04	0.17	0.04	0.39	0.26	0.09
5			0.1	0.2	0.4	0.3
6				0.33	0.33	0.33

Table 6.7: Conditional probabilities of stress given emotionality

$$H(y|x) = -\sum_{x=i}^{m_1} p_i(x) \sum_{y=j}^{m_2} p_{ij}(y|x) \log_2 p_{ij}(y|x) \tag{6.8}$$

It is the average sum of the entropies calculated on the conditional probability of the second variable given the first one.

Relationships with other entropies The relationship between conditional entropy and the other types of entropy is as follows:

$$H(Y|X) = H(X,Y) - H(X) \tag{6.9}$$
$$H(Y|X) = H(Y) - I(X,Y) \tag{6.10}$$

Conditional entropy of Y given X is the joint entropy of X and Y *minus* the entropy of X; in other words, it is the difference between the information of X and Y, and the information brought by X. This is clearly the meaning of this type of entropy, a reduction of uncertainty. The second variant $H(Y|X) = H(Y) - I(X,Y)$ puts it a little differently: it is difference between the information contained in Y and the shared part of X and Y.

Contrary to the joint entropy or the mutual information, conditional entropy is not a symmetrical measure: $H(X|Y) \neq H(Y|X)$. Conditioning on a variable or the other does not give the same result. This property will be fully exploited later, by conditioning the state of a system at time t *given* its state at time $t-1$. Of course it would not be the same as conditioning the other way around.

Example As an example we compute next the conditional entropies for the various combinations of the three binary variables, BEMOT, BSTRE and BFAM(see table 6.8).

We infer from the table that the best predictions relate to the knowledge of the familiarity of situation: knowing the state of BFAM $H(BSTRE \mid BFAM) = 0.69$ and $H(BEMOT \mid BFAM) = 0.54$. Meanwhile the worst prediction concerns the knowledge of the stress level: $H(BFAM \mid BSTRE) = 0.33$ and $H(BEMOT \mid BSTRE) = 0.39$. We may further notice the asymmetry of this measure by observing that $H(BSTRE \mid BFAM) \neq H(BFAM \mid BSTRE)$.

Variables	Conditional entropy H
H(BSTRE ∣ BEMOT)	0.56
H(BSTRE ∣ BFAM)	0.69
H(BEMOT ∣ BFAM)	0.54
H(BEMOT ∣ BSTRE)	0.39
H(BFAM ∣ BEMOT)	0.36
H(BFAM ∣ BSTRE)	0.33

Table 6.8: Conditional entropy between BEMOT, BSTRE and BFAM

Summary As a summary we may express the relations between the different entropies by the following Venn diagram (figure 6.4):

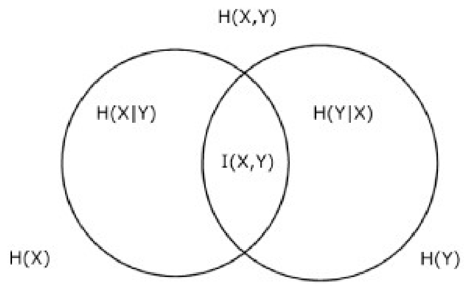

Figure 6.4: Relations between the different entropies

Each variable has its own *entropy* H(X) and H(Y). The shared part, the *mutual information* , lies at the intersection of the two variables. The *conditional information* is the information particular to a variable, while the *joint entropy* is the sum of the information of two variables.

The diagram makes clear that when the two variables are independent, the mutual information is 0, the conditional entropies equal the individual entropies and the joint entropy is the direct sum of the individual entropies.

Therefore the best measure of the proximity between variables is the mutual information. It was shown by Attneave that mutual information is related to the likelihood ratio G^2 by the following:

$$G^2 = 2N \ln 2 * I(X, Y) \tag{6.11}$$

where N is the number of observations, and ln(2), the natural logarithm of 2.

This is an important fact, since it is the link between the information theory and the statistical use of the mathematical branch (Bavaud, 1998).

6.1.5 Relative entropy (Kullback-Leibler distance)

Mutual information is a measure of the shared information between two variables. The *greater* the mutual information, the more similar two variables. Here we present a last measure that assess the opposite of the mutual information: the relative entropy, also known as the Kullback-Leibler's distance or the divergence.

Method Suppose two variables of the same type characterized by their probability distribution f and f'. The relative entropy formula is given by:

$$K(f||f') = \sum_{i=1}^{m} f_i \times ln\frac{f_i}{f'_i} \quad (6.12)$$

where m is the number of levels of the variables.

The properties of the relative entropy equation makes it non-negative and it is 0 if both distribution are equivalent ($f = f'$). The *smaller* the relative entropy, the more similar the distribution of the two variables, and conversely. It has to be noted that the measure is asymmetrical, the distance K(f || f') is not equal to K(f' || f). If the distributions are not too dissimilar, the difference between K(f || f') and K(f' || f) is small, and the distance is then equivalent to the chi-square statistics (relatively to the sample size n (Bavaud, 1998).

Example Suppose we would like to know if the stress experience of a subject has a close resemblance to her emotionality experience. In other words, we would like to answer the question: did she rated her stress level the same as the emotionality variable. May these two variables assess the same psychological concept, may be not. The relative entropy is an appropriate measure of the similarity of the underlying distribution.

Let's first compute the probability distribution of these two variables (table 6.9).

Variables	1	2	3	4	5	6
STRE (full-scale)	0.13	0.33	0.32	0.13	0.05	0.03
EMOT (full-scale)	0.09	0.54	0.23	0.09	0.04	0.01

Table 6.9: Probability distribution of STRE and EMOT

Replacing the probabilities in the Kullback-Leibler equation (6.12), we get [6]:

$$K(f_{STRE}\|f_{EMOT}) = 0.13 * ln(\frac{0.13}{0.09}) + 0.33 * ln(\frac{0.33}{0.54}) + 0.32 * ln(\frac{0.32}{0.23}) + ... = 0.100$$
$$K(f_{EMOT}\|f_{STRE}) = 0.09 * ln(\frac{0.09}{0.12}) + 0.54 * ln(\frac{0.54}{0.33}) + 0.23 * ln(\frac{0.23}{0.32}) + ... = 0.096$$

which implies that these two variables have a similar distribution. Note the slight asymmetry in the computation.

[6]Produced with the S-Plus command:

```
> RelativeEntropy(STRE,EMOT) and RelativeEntropy(EMOT,STRE)
```

Is the stress variable also similar to the other bio-psycho-social variables? Let's find out, by computing the relative entropy between STRE and the 7 variables. Figure 6.5 plots the resulting computations.

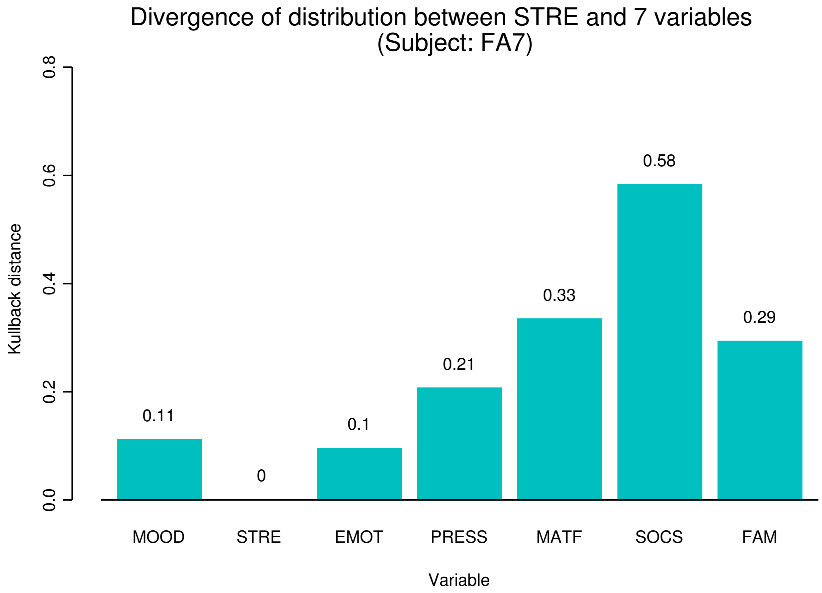

Figure 6.5: Relative entropy between STRE and 7 variables

As expected, we see that the minimal relative entropy (K=0) is between STRE and itself, simply because this entropy is 0 when applied to itself. The second and third lowest relative entropies are found for the internal variables EMOT and MOOD, while the highest are for the external variables PRESS, MATF, SOCS and FAM, social support being the most different.

The same procedure could be applied to assess the similitude between the distribution of a variable and the corresponding uniform distribution, or from any other distribution of interest.

Discussion Researchers must remember to eventually recode their variables before computing the distance. In our research the MOOD variable was rated as positive for values 4,5,6, and should be related to low stress, rated as 1,2 and 3; if we had not recoded MOOD into a reversed scale, the distance between these two variables would have been greater.

6.1.6 Conclusion

Having laid the foundations of information theory, researchers are ready to apply the knowledge on the time series data. Section 6.3.1 demonstrates how can one assess the complexity of transitions using conditional entropy. The latter measure is also applied for determining the *order* of a Markov chain (order means in Markovian framework the higher lag influences, not the opposite of disorder), as in chapter 8. But before delving into dynamical matter, let's answer a last "structural" question: the information loss when reducing scales.

6.2 Information loss when reducing scales

Our proposed methodology for dealing with multivariate configurations heavily relies on the dichotomisation of variables. The dichotomization process is interesting because it reduces the potentially large scale of the original data into a more schematic, compact result. But its major drawback is the crude, radical reduction of data range. Is it a too radical procedure? Is it within acceptable limits?

The method for answering this question is... *entropy* ! Simply because entropy is a measure of the quantity of information contained in a variable. If we compare the entropy of the original scale with the entropy for the recoded variable, we will get how much information is lost in the process.

This simple operation even allows to compare different recoding schemes. The investigator may wonder if a dichotomisation or a trichotomisation would be more efficient, or if dichotomisation around the mean or the median would be better. The entropy comparison yields a sound guideline.

Method We suggest computing the *relative loss of information* (or RLI), using equation 6.13:

$$RLI = \frac{H_s(\text{original scale}) - H_s(\text{new scale})}{H_s(\text{original scale})} \tag{6.13}$$

It is simply the difference between the standardized entropy of the original scale and the standardized entropy of the new scale, divided by the standardized entropy of the original scale, so to get a percent.

Example Let's return to our stress example. Recall from section 6.1.1 that the entropy of the full scale data is $H(STRE) = 2.20$ bits. Its standardized entropy is 0.85 (the standardized entropy is the entropy H (here 2.20) divided by the maximum entropy H_{max} ($H_{max} = \log_2 m = \log_2 6$).

The entropy of the new recoded scale STRE3 is $H(STRE3) = 1.33$ bits; its standardized entropy is $H_s = 0.84$ (computed by $\frac{1.33}{\log_2 3}$). This reduction of information from the original six-point scale to the new three-point scale amounts to 0.87 bits. On a standardized level, entropies are very close, 0.85 as compared to 0.84. The scale reduction seems to retain most of the original scale information.

Is this reduction of information a small or great quantity? The relative loss of information helps to determine it. The *RLI* of the trichotomisation equals to a 1.25% decrease of information. This is almost ridiculous, compared to the removal of 3 units in the original scale. We conclude that trichotomisation retains all information of the original scale while presenting a more compact scale.

What about the dichotomisation of the scale? This operation (as presented earlier) brings an entropy of $H(BSTRE) = 0.76$ bits, and a corresponding standardized entropy of 0.76 (indeed the same, since the maximum entropy of binary data is 1). There is a more or less important reduction of information here. The relative loss of information yields $RLI = 0.1107$, or 11.07%. This is also quite satisfying, since in the process 4 units in the original scale were removed, with a reasonable amount of loss. So we conclude that analyses based on binary variables may be continued.

What about other dichotomisation methods? We tried with a dichotomisation around the mean, which is 2.74, and around the median (which is 3). So the result is the same as with the center of the scale. We tried a dichotomisation at STRE=2; the entropy for this dichotomisation is $H = 0.995$, which is higher than the standardized entropy of the original scale! Recoding the variable this way must then be avoided. We also tried with a dichotomisation at STRE=4; results show that the entropy for this dichotomisation is $H = 0.42$, leading to a relative loss of information of 51.01%, a rather steep reduction.

Table 6.2 summarizes the findings.

	H_6	H_3	H_2 (center=3)	H_2(2)	H_2(4)
Entropy	2.20	1.33	0.76	0.995	0.42
Std Entropy	0.85	0.84	0.76	0.995	0.42
Reduction (bits)	-	0.87	1.44	1.20	1.78
RLI [7] (%)	-	1.25%	11.07%	$\uparrow 16.89$	51.01%

Table 6.10: Reduction of information from full scale to 3 and 2 value scales

As a further illustration of the procedure, the relative loss of information was computed for all 7 variables of our subject FA7. The computation is intended to compare the loss of information when dichotomizing the full-scale variables at the center of the scale, which is 3 for all of them. Results are shown on figure 6.6. They indicate that the variables for which the information loss is the greatest are familiarity of situations (FAM) and mood (MOOD). There even are two variables for which there is an *increase* of information, material support (MATF) and external pressure (PRESS); for these last two variables the dichotomization at the center of the scale should be avoided.

Discussion We have suggested here a way to compute the relative loss of information from one scale to another. It provides a sound guideline for deciding which scale is more appropriate. However it might be noted there still remains an arbitrariness in the procedure; there is an inherent compromise between the reduction of information and the range of scales that investigators want to keep.

Conclusion Recoding the original variable scale into a more compact one is suggested in this section. Three procedures are explained for the dichotomisation of variables: around a natural point if the scale is dichotomized; around the mean; around the median. A proposed formula, *relative loss of information* , computes the amount of information lost by such process. An example showed that while information is loss, some reduction of scale may be beneficial, retaining most of the original information in a more compact scale.

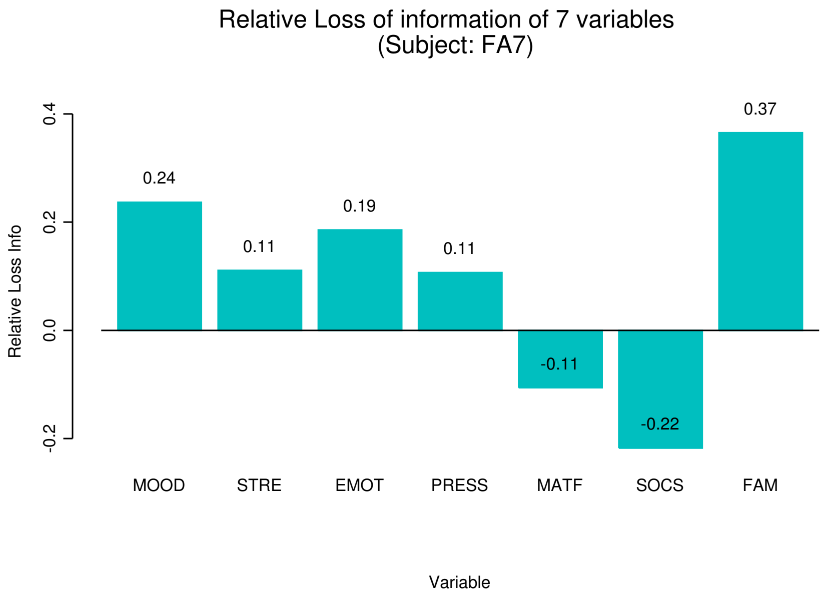

Figure 6.6: Relative loss of information of 7 bio-psycho-social variables

6.3 Measures of complexity of dynamics

Over the years the concept of complexity have been increasingly popular among scientific and non-scientific people. It has been used and abused. Even among the most rigorous scientists - mathematicians and physicists, there is no consensus on how to *compute* the complexity of a system.

Horgan (Horgan, 1995) made an inventory of no less than 30 different ways of defining and computing complexity. Among them there are:

1. algorithmic complexity
2. thermodynamics complexity
3. computational complexity
4. entropic complexity
5. grammatical complexity

Not surprising then it is regarded as a melting-pot.

However the fact that there are so many ways of defining and computing complexity should - and should not- scare us. Researchers have to be clear about their research objective, the nature of the data at hand and then proceed to select the appropriate measure of complexity.

Because of James Gleick's book on chaos, we were tempted in trying to detect presence of chaos in our data. Wouldn't have been nice to demonstrate that daily experiences are governed by nonlinear equations? represented by a chaotic attractor? that they follow fractals pattern?

We were thrilled by the possible discoveries, yet a minimum of realism was necessary. The data at hand were not of a sufficient quality (self-observations are less precise than hard psychometric and physiological measurements) and lack a sufficient variability (a 6-point Likert-scale does not provide a wide enough range) to make analyses for chaos.

6.3.1 An index of complexity of transitions

Information theory provides numerous examples of how to compute the complexity of a system, entropy being the best known case. Here we are interested in computing the complexity of the dynamics, or more specifically the complexity of the transitions. Intuitively following the entropic reasoning, we found a well-tailored measure of the *entropy of the transitions* , given the knowledge of the past state.

One intuition was to compute the "joint entropy" of a system, based on its transitional probability matrix. However, some states being more frequent than others, transitional probabilities would have the same weight. The entropy should be scaled so as to account for more important states than other and not be too much influenced by less frequent states.

Our expectations were answered by what scientists called the "entropy of an automorphism", or the "conditional entropy of order k". It was described in section 6.1.4 under the simpler name of conditional entropy. But we felt that a more appropriate name was needed, reflecting that this computation is performed on a dynamical perspective, not a synchronical one. So we nicknamed it the *index of complexity of transitions* .

Method Computing this index of complexity of transitions is done with this equation:

$$I_{ct} = -\sum_{ij}^{N} \pi_i * p_{ij} \log_2 p_{ij} \tag{6.14}$$

where π_i is the relative frequency of state i (number of states i divided by total number of states observed) and p_{ij} is the conditional probability of transiting from state i to state j. The index ranges from 0 to $\log_2 m$, where m is the number of possible different states; thus if there are 2, 8 or 16 states, the maximum index will be respectively 1, 3 or 4.

Equation 6.14 is exactly the same as equation 6.8; the difference between them is that here the conditional probabilities are transitional probabilities, which is a "subset" of all possible conditional probabilities.

Interpretation How does one interpret the index? The index mainly indicates the dispersion of the transitions, and thus the determinism of the process: with an index of 0, the process is completely deterministic, there is one and only one state j that follows state i. With a maximal index ($\log_2 m$), the number of states following each state i is maximal; transitions are then equiprobable. Following the reasoning, we may also say it represents the predictability of the system: a low index indicates a predictable system, and a high index, an unpredictable one.

Other interpretations of the index are possible. In connection with information theory interpretations, the smaller the index the less information and surprise is brought by the next state (and vice-versa, a greater index means that the next state brings more information). A deterministic system ($I_{ct} = 0$) is less informative regarding its transitions (however complicated the system is) than a random one ($I_{ct} = \log_2 m$) where each transition comes as a surprise.

Example As for our subject example (FA7), we computed the index of complexity of transitions for the three binary variables and its configural counterpart. Results are presented in table 6.11 [8]. Results show that the less complex dynamics is observed in the variable "familiarity of situation", with an index of 0.38; it is thus the variable for which predictions will be the most successful. The stress variable shows the highest index of complexity among the dichotomous variables, emotionality being in between. The complexity of the configuration is small at $I_{ct} = 1.28$, considering the maximum index is 3. The resulting dynamics is then quite predictable, although never perfectly.

Variable	Index of complexity
BSTRE	0.71
BEMOT	0.51
BFAM	0.38
Config	1.28

Table 6.11: Index of complexity of transitions for BFAM, BEMOT, BSTRE and Config

What about the other variables? We explore the question by computing the index of complexity of transitions on the 7 full-scale bio-psycho-social variables of the same subject (figure 6.7). On the whole less complex variables is given by two "external" variables, social support (SOCS) and familiarity of situation (FAM), while the most complex is the stress variables (STRE). It globally indicates that this subject's internal experiences followed a more complex path that the external conditions in which she lived in (exception being the material support, MATF).

As a matter of comparison we now examine the index of complexity of transitions of configurations from the other subjects who participated in our ESM study. We wonder if their dynamics is as structured as our subject example. Figure 6.8 graphically reports the findings.

The subjects' index of complexity of transitions ranges from 0.67 to 2.05, which spans the middle of the range of the index possibilities. We might then say that the underlying dynamics of these subjects is neither deterministic nor completely random. Our subject-example FA7 exhibits the third lowest index of complexity of transitions of this group.

[8] Produced with the S-Plus command:

```
> Index(BFAM), Index(BEMOT)...
```

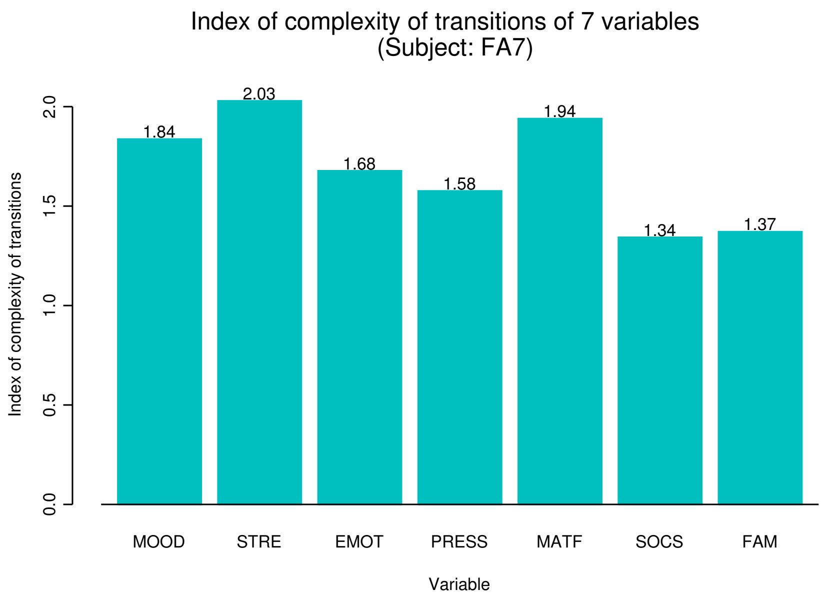

Figure 6.7: Index of complexity of transitions for the 7 bio-psycho-social variables

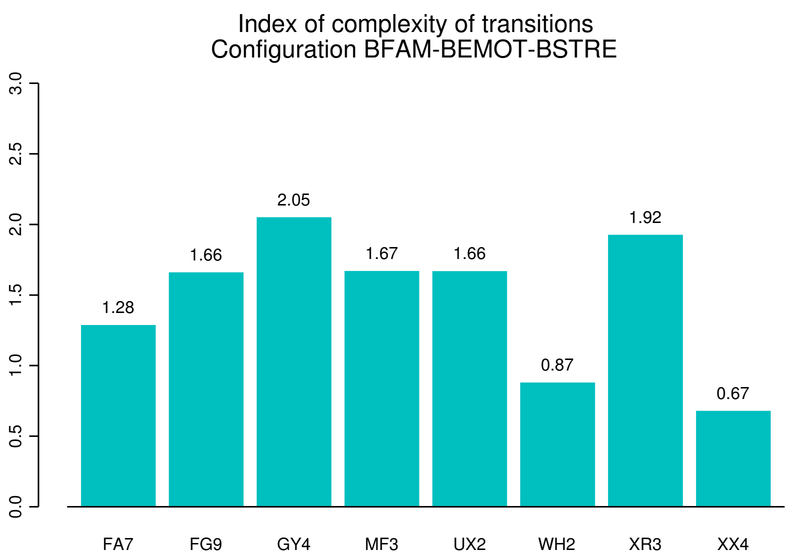

Figure 6.8: Index of complexity of transitions for the BFAM-BEMOT-BSTRE configuration (8 subjects)

Other uses of the index The index of complexity of transitions may be creatively used in other settings. Every time a transitional matrix is computed the index of complexity of transitions may be computed as well. Therefore researchers can easily compute the index on higher order transitional matrix (cf section 8.4). As with X^2-statistics "cross-transitional matrices" may be computed, that is, where the antecedent is the state of a variable X at time t and the consequent is the state of a variable Y at time $t+1$. Moreover one could examine the "local" index of complexity of transitions, that is, the index for each ij, much like what was done the analysis of residuals in transitional matrices; it would then inform on the "weight", or influence of each transitional probability on the index.

Discussion As many of the previous techniques discussed earlier, the numeric index alone does not indicate the nature of the process involved. Two equivalent indexes may refer to two very different dynamics. For example, an index of 0.15, meaning a quasi-deterministic process, could refer in one case to a subject who maintains a low level of stress and in another case to a subject who oscillates between high and low levels. These are two different processes measured by the same index. As previously advised researchers should always graphically represent the observed transitions using a state transition diagram.

A definite advantage of the index is that no parametric assumptions are made on the distribution of the variable; no normality is implied in the measure. Should transitions be concentrated into one category or many, occurred only once or as much as the other categories, all these situations are acceptable. The measure can be applied on categorical variables, and hence on configurations.

Its main limitation is that it is a descriptive measure only. No inferences are directly allowed. However it could be achieved using modern inferential techniques such as the surrogate technique.

Conclusion The index of complexity of transitions is definitively an appropriate measure for determining the amount of disorder in a dynamical system. Because of its sound theoretical framework (information theory), it should find its place in as computational arsenal of every researcher. And as stated earlier the index may be computed on higher order transitions; this is presented in section 8.4.

6.3.2 Other complexity measures

Rapp et all (Rapp, Jiménez-Montano, Langs, Thomson, & Mees, 1991) discussed a measure of grammatical complexity. Their goal was to give quantitative characterization of communication between patients and their therapist. The measure is said to be an upper bound of complexity, not the real complexity value, because it is not sure that this is the minimum description of a sequence.

Method The procedure is the following, starting with segment of length $n = 2$:

1. search the most frequent n-tuple in the sequence;
2. replace the most frequent n-tuple by a new symbol;

3. increase the length of the n-tuple by 1;
4. repeat steps 1 to 4 until no replacements can be performed;

When the original sequence has been "compressed" into another sequence of symbols, the complexity is computed as follows. It is the sum of the length of the new sequence plus the length of all symbols used to recode the sequence, without counting the repetition of symbols. If symbols are repeated in recoding, their logarithm is added to the complexity value.

The problem with this procedure relates to the lack of bounds or comparisons of the measure. One doesn't really know what a complexity of 5 or 20 means. Is it small or large? What does a disordered system typically exhibits? It makes the measure difficult to apply in real situations [9].

Another limit concerns the algorithm used to compute complexity. Starting with pairs, and after moving to triplets, quadruples and so on, is logical, but what if a system exhibits a 3-period cycle? Replacing pairs before triplets completely destroys the structure of the original sequence. Replacing triplets would be a far more efficient procedure.

In fact, an hybrid procedure, combining this approach with entropy would be better. One could compute the standardized entropy of pairs, triplets, quadruples and so on, and *then* replace the n-tuple that has the lowest standardized entropy, and repeat this operation on the recoded sequence until no other recoding is possible.

For these reasons we decided not to treat them in full detail here.

6.4 Advantages of information theory

Over the years information has developed a sound body of theories and statistical procedures. It is not only applied in engineering, but also in biological science, linguistic, and psychology.

As a whole information theory provides the necessary foundations for the statistical analysis of categorical variables. It may be used to characterize single variables (entropy) as well as groups of variables (joint entropy, mutual information, conditional entropy), on descriptive and inferential levels (by its relationships with G^2-statistics).

A major advantage of information theory is its *nonparametric* nature. Entropy does not require any assumptions about the distribution of variables.

Another advantage is that it does not assume a linear model (Golden, Brockett, & Zimmer, 1990). Our investigation was greatly determined by this constraint: we sought models for the analysis of dynamical systems that were not linear. Information theory answered this prerogative.

This assumption is important for categorical variables, but could also apply on quantitative variables. Suppose the following contingency table (cf. table 6.12) between two variables, where X represent any frequency.

Clearly there is a perfect relationship between the two variables. Belonging to a category of one variable implies the belonging to a specific category of the other. Information theo-

[9] By the time this thesis was written the author have been told a solution was found about this upper bound problem. But no article had been found due to particular circumstances.

	V2				
V1	1	2	3	4	5
1	0	X	0	0	0
2	X	0	0	0	0
3	0	0	0	0	X
4	0	0	0	X	0
5	0	0	X	0	0

Table 6.12: Hypothetical contingency table between two variables

retical analyses would detect this one-to-one relationship. Linear models would compute a zero Pearson correlation coefficient.

Another advantage relates to its applicability on categorical time series data. It will serve as the basis for computing dependences between states at various lags. There are a multitude of developments in mathematics and physics that seek to analyze the complexity of dynamical systems, via entropy-based computations. They will be described in later sections.

6.5 Limits of information theory

Information theoretical analyses consider variables as nominal ones. On ordinal variables the order of the scale is not preserved. So any analyses dealing with truly quantified variables may suffer a loss of information and power.

Another point, similar to the last one, is a loss of meaning in variables. In linear models, when for example correlation is computed, the direction of the relationship is known; if the sign is positive the two variables vary together, and when negative they do in opposite direction. Analyses on categorical variables seldom show the direction; a high mutual information does not tell for which categories there is a strong association. But this is a problem of categorical variables in general, not just the information theoretical approach.

6.6 Conclusion

Information theory is an encompassing framework for the analysis of categorical data. From the univariate measure of information using Shannon's entropy to its bi- and multivariate counterparts using joint entropy, conditional entropy, mutual information or relative entropy, researchers have in their hand a complete set of methods for dealing with this type of data. Care was taken to describe them with in details, current statistical handbooks being mostly silent about them. These methods were first examined on a synchronical level, and then on dynamical level, the index of complexity of transitions being a special case of the conditional entropy. We will later see how this index may be employed for assessing the influences of past states upon the present one (higher order Markov processes), and also for exploring the presence of a differentiated dynamics across time. This is the subject of the next chapter.

Please follow the guide...

Chapter 7

Analyzing phases

Summary. *When performing time series analysis, the sequence of data being examined must be stationary, that is, the parameters describing the dynamics must remain stable across the period of observation. Beside this methodological requirement, researchers may want to test the hypothesis that before and after a given event or moment, the dynamics either stayed the same or changed. They may also want to seek in a bottom-up approach those famous "bifurcation points", the moments that trigger radical change in the course of evolution of a system. This chapter is intended to answer these broad questions.*

Piaget described the growth of child cognitive capacities as a step by step progression, in a series of "plateaux", followed by abrupt bifurcation towards a higher order and degree of sophistication. The experiences of young individuals differ from those an older age. The daily routine of individuals during weekdays is typically transformed during weekends.

These are illustrations of a change in dynamics that occurs across time. When examining the dynamics of a system during a period of time and then another one sees that the structure in the sequence of observations is not the same. There appears to be *phases* , that is, periods of change.

The analysis of the *dynamics of dynamics* , e.g. the stability or change of dynamics over time is in fact a methodological requirement. All the previous analyses presuppose that the dynamics does not change over time. The sequence of data needs to have the same structure in the transitions; it needs to exhibit *stationarity* .

7.1 What is stationarity?

Technically a time series is said to be stationary if the distribution of a variable $X_1, X_2, \ldots, X_n$ is the same as the distribution of the variable shifted by some lag k, $X_{1+k}, X_{2+k}, \ldots, X_{n+k}$; the distribution of the variable does not depend on time t. Wherever one looks at the distribution for some segment the dynamics remains the same. This means that it does not matter when in time we observe the process.

A convenient but weak definition of stationarity regarding quantitative variables is that there is no systematic change in either mean or variance in the time series. If there were such changes an increasing or decreasing trend in the data would be present. But this is somewhat a weak definition since a constant mean over time may result from very different dynamics.

7.2 Testing stationarity

Is the dynamics of a system constant over time? Are transitions between two states i and j governed by the same "law" or mechanism over time? We will provide the necessary background for answering these questions.

Method An appropriate method, called "omnibus", relies on the χ^2 statistic (Gottman & Kumar Roy, 1990). What needs to be done is to divide the whole sequence of observations (of length n) into D time periods, thus yielding D subsequences (of length n/D. The transitional frequency matrix for each time period is then computed; a statistical test will compare the individual transitional frequency matrices with the overall one.

Let $n_{ij}(d)$ be the transitional frequency from state i to state j during the time period d. Let there be D such time periods. There are then D transitional frequency matrices, organized in the following fashion (cf table 7.1).

The transitional probabilities are also computed for each time period as well as their expected values. $p_{ij}(d)$ is the conditional probability of the transition from state i to state j during time period d, and p_{ij} (without subscript) is the conditional probability of the whole sequence (the sum of all transitional frequencies $n_{ij}(d = 1 \ldots D)$ divided by the marginal frequency n_{i+}. It is thus estimated by $\frac{n_{ij}}{n_{i+}}$.

The estimated expected transitional frequency in the d-the time period is given by the following equation, where $n_{i+}(d)$ is the marginal frequency of state i during time period d:

$$e_{ij}(d) = n_{i+}(d)\frac{n_{ij}}{n_{i+}} \tag{7.1}$$

Having both the observed and the expected transitional frequencies, we are ready to compute the X^2 statistics, using:

$$X^2 = \sum_{d=1}^{D}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}(d) - e_{ij}(d))^2}{e_{ij}(d)} \tag{7.2}$$

which is asymptotically distributed as χ^2 with $D(J-1)(I-1)$ degrees of freedom.

One may also use the similarly distributed likelihood ratio statistics G^2, using this equation:

$$G^2 = 2\sum_{d=1}^{D}\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(d)\log_e \frac{n_{ij}(d)}{e_{ij}} \tag{7.3}$$

	$X(t+1)(1)$				
	1	2	$\dots$	m	total
$X(t)(1)$ = 1	$n_{11}(1)$	$n_{12}(1)$	$\dots$	$n_{1m}(1)$	$n_{1+}(1)$
2	$n_{21}(1)$	$n_{22}(1)$	$\dots$	$n_{2m}(1)$	$n_{2+}(1)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
m	$n_{m1}(1)$	$n_{m2}(1)$	$\dots$	$n_{mm}(1)$	$n_{m+}(1)$
total	$n_{+1}(1)$	$n_{+2}(1)$	$\dots$	$n_{+m}(1)$	$n_{++}(1)$

	$X(t+1)(2)$				
	1	2	$\dots$	m	total
$X(t)(2)$ = 1	$n_{11}(2)$	$n_{12}(2)$	$\dots$	$n_{1m}(2)$	$n_{1+}(2)$
2	$n_{21}(2)$	$n_{22}(2)$	$\dots$	$n_{2m}(2)$	$n_{2+}(2)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
m	$n_{m1}(2)$	$n_{m2}(2)$	$\dots\dots$	$n_{mm}(2)$	$n_{m+}(2)$
total	$n_{+1}(2)$	$n_{+2}(2)$	$\dots$	$n_{+m}(2)$	$n_{++}(2)$

$\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots$

	$X(t+1)(D)$				
	1	2	$\dots$	m	total
$X(t)(D)$ = 1	$n_{11}(D)$	$n_{12}(D)$	$\dots$	$n_{1m}(D)$	$n_{1+}(D)$
2	$n_{21}(D)$	$n_{22}(D)$	$\dots$	$n_{2m}(D)$	$n_{2+}(D)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
m	$n_{m1}(D)$	$n_{m2}(D)$	$\dots$	$n_{mm}(D)$	$n_{m+}(D)$
total	$n_{+1}(D)$	$n_{+2}(D)$	$\dots$	$n_{+m}(D)$	$n_{++}(D)$

Table 7.1: Structure of transitional frequency matrices in D time periods

Hypothesis The tested hypothesis is that the transitional probabilities are constant across time periods. It is mathematically expressed as:

$$H_0 : p_{ij}(d) = p_{ij}, \text{for } d = 1, 2, \ldots, D$$
$$H_1 : p_{ij}(d) \neq p_{ij}, \text{for } d = 1, 2, \ldots, D$$

It states that transitional probabilities of each time periods should be equal to the transitional probabilities of the whole sequence.

Interpretation If the null hypothesis is accepted, we conclude that there is no change in the dynamics across time. It is stationary, or stable. If H_0 is rejected we conclude that the transitional probabilities are different across time periods, that is, the dynamics changed across time.

Example As an example, we will examine the stationarity of the dichotomized stress variable BSTRE. In particular we will test whether there is a change in the dynamics at mid-term, e.g. before and after the middle of the observation period. There are then two transitional frequency matrices of length 131 and 130 to be compared with the global transitional frequency matrix of the whole sequence (length 261).

On table 7.2 are represented both the transitional frequencies and probabilities of the first and second time periods, and two resulting global matrices.

The X^2 and G^2 tests for stationarity were performed, resulting in the table 7.4 [1].

Results show a non-significant test, the two p-values being well above the usual $p = 0.05$. That is, the transitional probabilities of the two time periods should be considered equal. We then accept the hypothesis that the dynamics underlying the two time periods are similar, hence the hypothesis of a stationary dynamics. That implies all analyses performed on the BSTRE global transitional frequency matrix are statistically valid.

What about the two other variables, emotionality and familiarity of situations? Here are the results of the omnibus test on these variables (using the X^2 statistic; cf table 7.5) [2].

Interestingly the omnibus test on the BEMOT variable is significant, but not for the BFAM variable. In the first case it implies that before and after the 131-the observation, the dynamics of emotionality changed; the transitions are not equivalent across the two time periods. The sequence of observations is therefore *non-stationary* , which impedes the other analyses performed on this variable. In the second case the dynamics before and after that point remains the same, yielding a stationary dynamics.

Examining the transitional frequencies, as in table 7.6 we readily see how the dynamics differ. There is a shift in the number of transitions between the first period and the second period related to low emotionality; at first there is a high number of transitions for $BEMOT_{0\to 0}$ ($n_{00} = 88$) and fewer for the three other transitions ($n_{01} = 14$, $n_{10} = 15$ and $n_{11} = 13$); these frequencies change afterwards, leading an even higher frequency for

[1]Produced with the S-Plus command:

```
> Omnibus(BSTRE)
```

[2]Produced with the S-Plus command:

```
> Omnibus(BEMOT) and Omnibus(BFAM)
```

Transitional Frequencies / **Transitional Probabilities**

1st time period

	BSTRE(t+1) 0	BSTRE(t+1) 1		BSTRE(t+1) 0	BSTRE(t+1) 1
BSTRE(t) = 0	90	15	BSTRE(t) = 0	0.86	0.14
1	16	9	1	0.64	0.36

2nd time period

	BSTRE(t+1) 0	BSTRE(t+1) 1		BSTRE(t+1) 0	BSTRE(t+1) 1
BSTRE(t) = 0	80	17	BSTRE(t) = 0	0.82	0.18
1	18	14	1	0.56	0.44

Global time period

	BSTRE(t+1) 0	BSTRE(t+1) 1		BSTRE(t+1) 0	BSTRE(t+1) 1
BSTRE(t) = 0	170	33	BSTRE(t) = 0	0.84	0.16
1	34	23	1	0.60	0.40

Table 7.2: Transitional frequencies and probabilities for two time periods (BSTRE)

Expected Frequencies $e_{ij}(d)$ — **Residuals**

1st time period

	BSTRE(t+1) 0	1
BSTRE(t) = 0	87.93	17.07
1	14.91	10.08

	BSTRE(t+1) 0	1
BSTRE(t) = 0	0.05	0.25
1	0.08	0.12

2nd time period

	BSTRE(t+1) 0	1
BSTRE(t) = 0	81.23	15.77
1	19.09	12.91

	BSTRE(t+1) 0	1
BSTRE(t) = 0	0.02	0.10
1	0.06	0.09

Table 7.3: Expected values and χ^2 residuals for two time periods (BSTRE)

Variable	splitting point	statistic	d.f.	p-value
χ^2	131	0.765	2	0.68
G^2	131	0.774	2	0.68

Table 7.4: Test for stationarity on BSTRE using χ^2 and G^2 statistics

Variable	splitting point	X^2	d.f.	p-value
BEMOT	131	7.49	2	0.02
BFAM	131	1.81	2	0.40

Table 7.5: Test for stationarity on BEMOT and BFAM using χ^2 statistic

1st time period

	BEMOT(t+1)	
	0	1
BEMOT(t) = 0	88	14
1	15	13

2nd time period

	BEMOT(t+1)	
	0	1
BEMOT(t) = 0	116	5
1	6	2

Table 7.6: Transitional frequencies for two time periods (BEMOT)

$BEMOT_{0\to0}$ ($n_{00} = 116$) and frequencies close to 0 for the other transitions. That means the second period was mostly characterized by low emotionality, while there were episodes of higher emotionality during the first half.

Example with more than two periods The procedure described previously also applies when testing for more than two time periods. Investigators could be interested in sectioning their sequence of observations in 3, 4 or more subsequences. Here we examine the dynamics of the three dichotomous variables (cf. table 7.7) [3]. We wonder if their dynamics is subject to finer distinctions than a crude half-half sectioning. The overall sequence is divided in three segments of length 87.

Variable	length	X^2	d.f.	p-value
BSTRE	87	3.77	3	0.29
BEMOT	87	14.39	3	0.002
BFAM	87	14.21	3	0.003

Table 7.7: Test for stationarity on BSTRE, BEMOT and BFAM using χ^2 statistic, for 3 subsequences

Results indicate that if the dynamics of the stress variable does not change by considering 3 subsequences, the dynamics of emotionality and familiarity of situations does change. Examining the transitional frequency matrices for these variables on each of the subsequences allow to detect the source of this modification. Indeed this subject encountered most of her unfamiliar situations during the middle of the observation period. Concerning the emotionality, the transitional frequency matrices show a continuous decrease of emotionality, leading to more and more frequent $BEMOT_{0\to0}$ transitions.

A question now naturally arises: what if more subsequences are taken into account? Why not considering 4, 5, ..., subsequences? Using computer power allows to iterative test for any number of subsequences. But there is nevertheless an important limitation: increasing a greater number of segments decreasing the length of these segments. Considering that the use of chi-square statistics on transitional frequency matrices becomes problematic when

[3]Produced with the S-Plus command:

```
> Omnibus(BSTRE, d=3)
```

there are less 80% of cells have a frequency less than 5, the minimal length of a sequence should be $5 \times m^2$, where m is the number of levels. Therefore the minimal sequence length for a binary variable is $5 * 2^2 = 20$, $5 * 3^2 = 45$ for a trichotomous variable, and so on.

Consequently, for testing our binary variables, given this minimal length of 20 observations, we could examine up to $261/20 = 13$ segments, but will consider no more than 10. Figure 7.1 [4] shows the results of this iterative process.

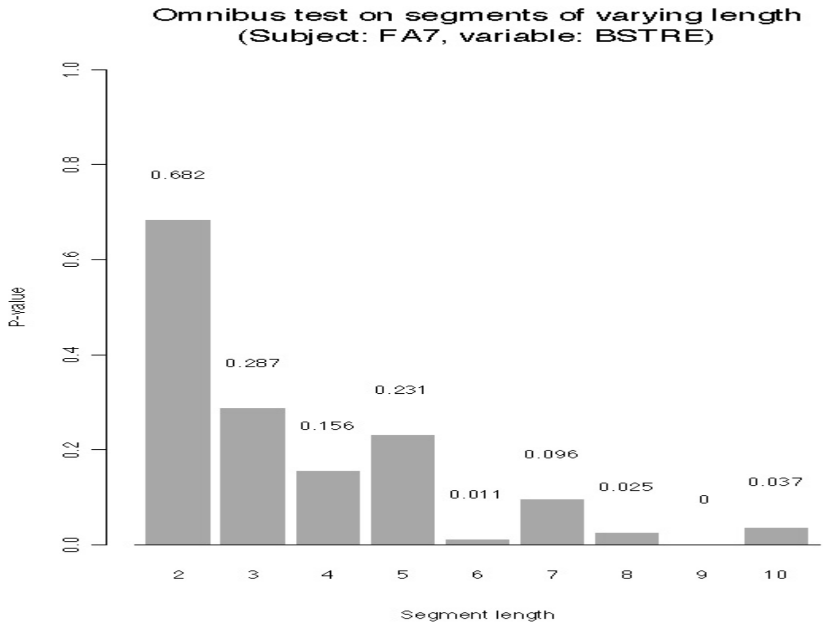

Figure 7.1: P-value of the omnibus test BSTRE on segments of length 1-10

As previously explained the multiple testing procedure requires to decrease the α-level at $0.05/9 = 0.0055$. Any test whose p-value is small than 0.0055 indicates that there is a differentiated dynamics across the observation period. For this stress variable, there is only 1 omnibus test for which the p-value is smaller than 0.006 and it is when the overall sequence is divided into 9 segments of length 29. There is expectedly a general decrease in the p-value when the number of segments are increased: this is explained by the corresponding decrease in the length of observations taken into account.

Discussion Following the line of contingency table analysis, the omnibus test allows investigators to test the stability or change in the dynamics. It corresponds to the analysis of the *dynamics of dynamics* .

We presented a test based on X^2 and likelihood ratio statistics, G^2. Because these two statistics usually show similar behavior, only results from the X^2 test were shown.

[4]Produced with the S-Plus command:

```
> for (i in 2:10) bstre.omnibus[i-1] _ Omnibus(BSTRE, d=i, matrices=F)$px
```

The methodology proposed here works for the comparison of two or more time blocks. When using this test one should pay attention to the number of time periods. Increasing the number of segments decreases the individual cell transitions, potentially leading to sparse tables and their associated problems.

Conclusion Investigators have a very good technique to test the stationarity of the data sequence.

The technique may be used in a *methodological perspective* , that is, for testing the stationarity itself. In this case the splitting point is chosen at half-point, so to separate the sequence in two equal parts.

It also be used in a truly experimental fashion, for testing the influence of a specific moment on the course of event. The event may be the result of a research protocol, where experimenters provoke a situation susceptible to change –or not– the future dynamics. The test then serves to determine if a significant change has occurred because of the event. In an exploratory study without a priori expectations the test may serve to test the influence of a naturally happening event.

7.3 Moving-omnibus: a bifurcation-point detector

A statistical test based on χ^2 and G^2 were presented for examining the change of dynamics before and after a certain moment. This moment is often chosen to be the point which separate the data sequence in two halves. But dividing at half-point may not be the best choice, or at least the most informative.

An event that occurred 10 minutes, 6 days or 2 months after the beginning of observation may radically transformed the dynamics of the system (depending on the time scale adopted). Complex dynamical systems theory calls these moments *bifurcation points* (Haken, 1983; Barton, 1994). Splitting a sequence at a single point has more chance *not* to capture the influence of this determining event than to do so.

We will describe a procedure which takes advantage of the power of computer, to systematically assess the influence of almost all events in the sequence.

Method The suggested procedure consists in computing the omnibus test not just for the half-point, but also for splitting points at n=30, n=65, n=120, ... Well, for all possible points that lead to sound statistical analysis.

What are the lowest and highest splitting point that researchers can start with? As explained before, the minimal length of a sequence is given by $5m^2$. So the lowest splitting point is defined by this limit, $5m^2$, Similarly the highest splitting point is defined as the length of the sequence minus $5m^2$ (that is, $n - 5m^2$).

The complete procedure resembles this pseudo-code snippet:

1. let n be the number of observations and m the number of modalities in the variable
2. from $n = 5m^2$ to $(n - 5m^2)$, compute the Omnibus test splitting the sequence at point n
3. plot the evolution of the test for each n

And because of the high number of tests involved and in order to reduce the risk of rejecting hypotheses on a basis of chance, researchers are encouraged to lower the α-level of the tests. Applying Bonferonni's correction, one should divide the α-level by the number of tests involved. For example, in our example there are 261 - 20 -20 = 221 omnibus tests performed; therefore α should be set at $\frac{0.05}{221} = 0.0002$

Example For our 2 state variable BSTRE, this *Moving-omnibus* procedure is employed. The first splitting point is 20, so the first comparison examines the transitional frequency matrix comprising data 1 to 20 with the transitional frequency matrix comprising data 21 to 261. The second comparison is between data sequence 1 to 21 with 22 to 261, the third goes from 1 to 22 and 23 to 261, and so on. The procedure increases the splitting point by one observation, thus increases the length of the first data sequence by one observation and decreasing the second data sequence by one.

The results are presented in the form of a graph of evolution (figure 7.2) [5]. The x-axis shows the value of the splitting point, and on the y-axis is shown the p-value of the test (of course the computation of the omnibus test yields the usual three parameters: χ^2 value, degree of freedom and p-value, only the p-value is shown here).

Figure 7.2: Evolution of stationarity test (χ^2) on BSTRE

The figure shows that never on the course of the sequence the dynamics radically changed to as to become statistically different. The figure tells us that the dynamics is not strictly stable across time. There are even points where it almost reaches significance: before and after point 25, 50 and 205.

For the STRE3 variable, which was shown to be a non-stationary sequence, the evolution of the significance is on figure 7.3 [6]. It clearly demonstrates that the period between n=115 to n=195 provides separating points where the dynamics before and after is different.

Figure 7.3: Evolution of stationarity test (χ^2) on STRE3

The evolution of the stationarity test for the STRE6 is shown in figure 7.4 [7]. Like the figure for the BSTRE variable, the dynamics never reaches significance, although there are changes in the dynamics around the middle of the observation periods. But as expressed earlier, tests with this variable is subject to caution, due to numerous low transitional frequencies.

[5] Produced with the S-Plus command:

```
> for (i in 2:10) Omnibus(BSTRE, d=i, matrices=F)$px
```

[6] Produced with the S-Plus command:

```
>  for (i in 2:10) Omnibus(STRE3, d=i, matrices=F)$px
```

[7] Produced with the S-Plus command:

```
>  for (i in 2:10) Omnibus(STRE,d=i, matrices=F)$px
```

Figure 7.4: Evolution of stationarity test (χ^2) on STRE6

Discussion The moving-omnibus technique answers a fundamental question: is the dynamics of the system stable? By iteratively computing the omnibus test for all possible (reasonable) splitting point of a sequence, researchers are able to point out moments where the system bifurcated into a significantly different type of dynamics. Even for researchers who had a priori hypotheses about the influence of one specific event, this bottom-up approach may reveal other pivotal moments in the course of events.

This automated procedure is very interesting. However as powerful as it is, it always requires the careful analysis of the investigator. It surely does not dispense him to examine the transitional frequency matrices, should the omnibus tests be significant or not. Significance may be explained by "causal" reasons, or by methodological bias (like small transitional frequencies).

7.4 Moving-entropy and moving-index of complexity

Detecting moments corresponding to bifurcations in the evolution of a system is perhaps the most important objective when analyzing the system dynamics in terms of phases. But there are also many others, that will be detailed here.

While dealing with naturally occurring systems such as daily individual experiences, we have noticed that people have typical routine patterns, that are sometimes disturbed by some external events acting as perturbations. The goal then was to identify those typical patterns of experiences and behavior, and see how people react to the "random" perturbations.

We mainly described the routines as attractors, defined by multivariate configurations of bio-psycho-social variables, where people tended to stay most of the time, and towards which they returned if some other perturbations occurred.

We have seen that most of our subjects were driven by the same attractors; they sought familiar situations, socially and materially supportive of their actions, for which they could experience low stress and emotionality, thus being in a positive mood. It seems that what differentiates the subjects is not the attractor in itself – it seems rather odd to think they would seek the opposite of these experiences – but the *variability* around this attractor.

This variability is not just a *global* feature, some subjects generally exhibiting a more or less variable number of different experiences. It is also a *local* , individual feature. Indeed there are phases where the subjects' experiences are *crystallized* around the same type of experiences, described by the attractor. And during other phases they undergo many changes, fluctuating and oscillating in an important fashion around this attractor.

We describe in this section procedures which again takes advantage of computer power, to describe the *evolution of the variability* of systems. Researchers may then spot the phases when systems are crystallized in a routine pattern, and phases where fluctuations where very important.

Method The main idea behind the following procedures is to compute a certain statistic for a time window of a certain length, and then move the time window by one observation, until computations are no longer possible.

The procedure resembles this pseudo-code snippet:

1. let N be the number of observations and l, the length of a time-window
2. from n = 1 to $(N - l + 1)$, compute the desired statistic
3. plot the evolution of the statistic for each n

The length of the time-window must be chosen on pragmatical ground, both from an empirical and statistical point of view. For our research, since subjects yielded 7 self-observations each day, a time window of 49 observations (thus including one week) is an appropriate choice. Shorter sequences may be taken into account, but then statistical principal prevails; remember the discussion about transitional frequency matrices and chi-square statistics.

Contrary to the moving-omnibus procedure, which moves a splitting point across the sequence of observations, the next procedures move a complete time-window, or subsequence of length l.

We describe here two methods, one based on entropy, the other based on the index of complexity of transitions. What we want first is to determine the variability (or the dispersion, like the standard deviation) of the observations; the second method assess the complexity of the dynamics.

Example As previously describe we use here a time window of length 49, so to include periods of one week of observations. The time-window is thus moved from $n = 1$ to $n = 261 - 49 + 1 = 211$. Entropy and then index of complexity is computed for each of these time-windows.

Moving-entropy We begin with the evolution of the entropy (cf figure 7.5) [8]. The graphic represents the evolution of the three binary variables stress (BSTRE), emotionality (BEMOT) and familiarity of situation (BFAM), as well as the configuration formed by those variables (Configuration).

The graphic shows that the entropy of the system is not a static characteristic, but a dynamical one. It changes across time, decreasing, increasing and remaining stable for periods of time.

The moving-entropy of the configuration ranges from 0.75 to 2.0, but dispatched in phases. It first begin around 2.0, dropping to 1.0 reaching a plateau between point 25 to 90; it then rises again to 2.0 until point 140, dropping again down to 0.75.

Regarding the interplay between the entropies of the individual variables and the configuration, we infer from the graphic that the *level* of the configuration entropy mainly comes from the stress variable, while the *variation* of the configuration entropy is mainly contributed by the emotionality and familiarity of situation variables. On the one hand it appears that

[8] Produced with the S-Plus command:

```
>  for(i in 1:(261-49)) Entropy(BSTRE[i:(i+49-1)])
```

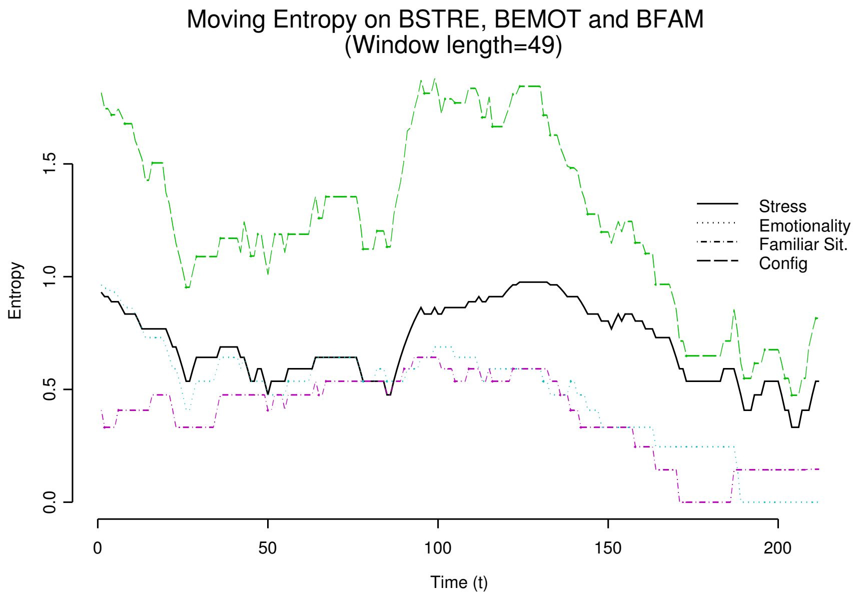

Figure 7.5: Evolution of the moving-entropy (window length=49)

the stress exhibits a higher entropy and is more stable across time. On the other hand the other variables generally exhibit a lower entropy level, and vary more, radically decreasing at the end of the observation period.

Moving-complexity The latter section presented the evolution of the entropy, so as to investigate the variability or dispersion of the variables. We now describe the procedure for analyzing the evolution of the complexity of the transitions. We employ of course the index of complexity of transitions as the basis for computation.

Figure 7.6 [9] shows the evolution of the index of complexity of transitions, using a 49 observation time-window.

The evolution of the index of complexity closely resembles the evolution of the entropy shown on the precedent figure. The index of the configuration ranges from 0.75 to 2.00; starting below the 1.50 level, it drops under 1.00 at point n=30, then slowly rises above 1.50 at n=130 to drop around 0.50 at the end of the sequence. On the one hand it implies that the dynamics of this subject's experiences was never completely crystallized into a fixed routine, although by the end it seemed to stabilize considerably. On the other hand it neither became very disorganized.

[9]Produced with the S-Plus command:

```
>  for(i in 1:(261-49)) Index(BSTRE[i:(i+49-1)])
```

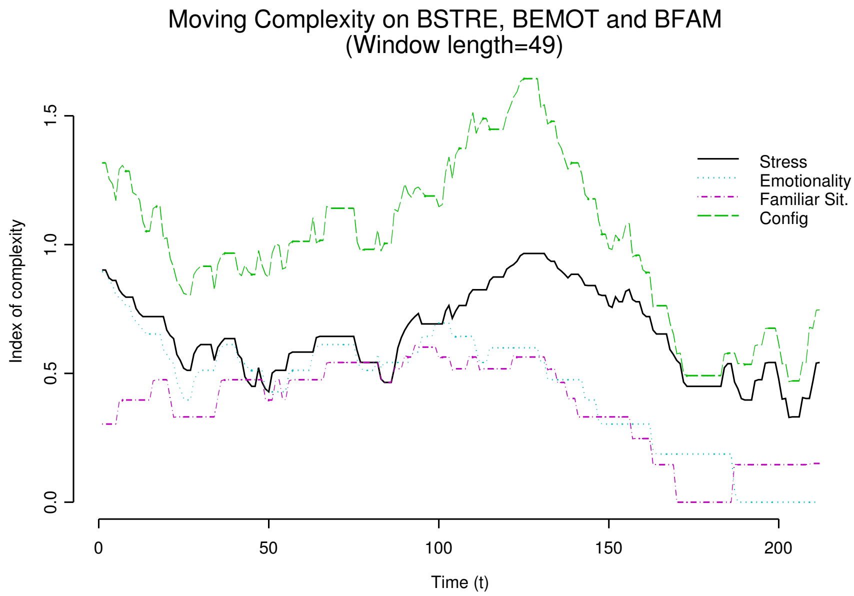

Figure 7.6: Evolution of moving-complexity (window length=49)

Moving-tests As one could easily imagined, these moving-test procedures are of sufficiently general nature to allow other variations. Among other fruitful perspectives there could be:

- a moving-chi-square procedure, where we could test the presence of a first-order Markov process;
- a moving-lambda procedure, to examine the amount of uncertainty reduction in sequence of given length;
- a moving-mutual-information procedure, for assessing the amount of shared information between two variables;
- a moving-Sackett-z procedure, for testing specific patterns of transitions;

The list could quite long, as any statistical test performed on a sequence of observation, given researchers fulfill the methodological requirements, may be transformed into a moving procedure. Researchers are encouraged to use their imagination and develop methods that will increase the knowledge about the underlying dynamics.

7.5 Conclusion

Principles and techniques for analyzing the presence of phases were described in this chapter. We first explained what is stationarity, why it is a methodological requirement. We proceeded to demonstrate how it is tested, using the omnibus test, based on familiar chi-square statistics. Testing stationarity may also be a research objective, for researchers may want to determine the influence of a given event on the course of observation. We then presented iterative procedure named moving-omnibus, moving-entropy and moving-index that allow to seek in a bottom-up fashion potential bifurcation points. Various other procedures could be developed, some were briefly outlined.

Chapter 8

Higher order Markov chains

Summary. *Most analyses performed on categorical time series data refer only to the dependence between the present state and its past state; it tests whether the process is first order Markov chain. But often researchers are interested in delving into past state influences. To what extent do past states determine the present state? How many lags should be examined when analyzing higher orders? This section is intended to answer these questions.*

A few methods exist for analyzing the order of Markov chain processes. Some are based on information theory. The rationale behind their reasoning is that taking into account past states should reduce the surprise or the uncertainty brought by the present state.

We first present methods based on auto-correlation coefficients and the index of complexity of transitions, for a more exploratory stance. These methods do not involve a statistical test *per se* (but auto-correlations significance may be assessed). Because there are no auto-correlation coefficient that can be applied on configurations, researchers must rely on chi-square statistics; the procedure is examined afterwards. Another presented method employs both the entropy and a statistical test to determine the order. We close this chapter by presenting methods that examine the order of specific transitions.

8.1 Higher order computations

Considering higher order influences implies computations of transitions differing from the first order transitional matrices. There are two methods for achieving this goal: the higher order transitional (frequency and probability) matrix and the n-grams method. We now examine how their are developed.

8.1.1 Higher order transitional matrices

Transitional matrices for higher order are build by considering the number of transitions from state i to j, k states before, *without* taking into account the states in-between, as would an n-gram procedure suggests.

	V1 $(t+1)=0$	$V1(t+1)=1$
$V1(t)=0$	0.00	1.00
1	0.46	0.54

	V1 $(t+2)=0$	$V1(t+2)=1$
$V1(t)=0$	0.00	1.00
1	0.50	0.50

	$V1(t+3)=0$	$V1(t+3)=1$
$V1(t)=0$	1.00	0.00
1	0.00	1.00

Table 8.1: Transitional frequency matrices for 2nd & 3rd order (third-order cycle)

Examples Let's first illustrate with a trivial but prototypical case: a deterministic sequence of third-order, given by the repeated sequence 1,1,0:

$$1,1,0,1,1,0,1,1,0,1,1,0,1,1,0,1,1,0,\ldots \tag{8.1}$$

When considering the first order ($k=1$), we take into account the strictly consecutive pairs $s_{i,i+1}$, which are (1,1), (1,0), (0,1), (1,1) and so on. But when considering higher order k, we take into account pairs $s_{i,i+k}$, leaving aside intermediary states. For a second-order $k=2$, those pairs $s_{i,i+2}$ are (1,0), (1,1), (0,1), (1,0), ... For a third-order $k=3$, those pairs $s_{i,i+3}$ are (1,1), (1,1), (0,0), (1,1), ... We may continue the procedure for high orders, sometimes up to $k=10$, a very high order for which the number of data usually is insufficient for computing significance with enough precision. Let's just show the transitional probability matrices for the first three orders (table 8.1).

The example clearly shows the data exhibits a completely deterministic third-order cycle, the transitional probabilities being 1.00 for the $0\to 0$ and $1\to 1$ transitions. If an investigator had examined only the first and second order, he would probably conclude that the dynamics converges towards a solution where after a 0 always follows a 1, and after a 1 there would be equal chances to return to 0 or stay in 1.

Now let's see if our dichotomized stress variable example behaves similarly. The 20 first observations are shown below:

$$1,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0,0,0,0,1,\ldots \tag{8.2}$$

For the first-order $k=1$, the transitional probability matrix have already been presented (in section 5.6). For the second-order $k=2$ and $k=3$ we have results in table 8.2.

It is difficult to infer a definite pattern considering only these matrices. We postpone their detailed analysis in the later section.

Discussion Question arises if the consideration of higher orders converges to a more predictable dynamics. Testing the significance and the complexity of these transitional matrices

	$BSTRE(t+2)=0$	$BSTRE(t+2)=1$
$BSTRE(t)=0$	0.80	0.20
1	0.74	0.26
	$BSTRE(t+3)=0$	$BSTRE(t+3)=1$
$BSTRE(t)=0$	0.82	0.18
1	0.66	0.34

Table 8.2: Transitional frequency matrices for 2nd & 3rd order (binary stress variable)

are latter examined. For the moment researchers should only be ware that the dynamics does not necessarily (although it may) converge to the limit matrix of Markov chains analysis (section 3.4); it will converge only if the dynamics is best summarized by a first-order Markov chains. Stated differently, the k-order transitional probability matrix is not necessarily equal to the first-order transitional probability matrix taken to the k-the power.

8.1.2 The n-gram method

The second approach to dealing with higher-order transitions is to compute n-grams. The idea resembles the building of synchronical configurations, but on a dynamical perspective.

Method Remember that configurations were built by appending together n variables (at same time t) to form a new single entity. Here instead of appending n synchronical variables, n-grams are formed by concatenating the last n states (of the same variable) into a single entity. But contrary to the last method, the n-gram method takes into account the intermediary states.

Example Let's see how n-grams are computed. Considering the first order ($k = 1$), all $n = 2$ possible combinations of consecutive states are concatenated into a single symbol; they are 00, 10, 01, 11. But when considering higher order k, we take into account pairs, integrating intermediary states. For a second-order $n = 3$-grams, the possible symbols are 000, 001, 010, 011, 100, 101, 110, 111.

Once the n-grams are formed their empirical frequencies and probabilities of occurrences are computed. On table 8.3 are shown the frequencies of the n-grams for the subject's binary stress variable, for n up to 4.

The table is read from right to left. For $n = 1$, the table informs that the symbol BSTRE=0 is found 204 times in the sequence, and BSTRE=1 is there 57 times. For $n = 2$, the 204 occurrences of the 0 symbols are the result of 170 transitions composed of 00, and 34 transitions of 01; similarly, the 57 occurrences of symbol 1 result from 33 transitions 10 and 23 transitions from 11. We may continue the reading up to the quartogram, the $n = 4$-gram.

Does taking into account a supplementary symbol in the past reduce the uncertainty about the present state? Information-based methods will statistically assess which $n$grams bring the less information, that is, those which provides the most compact representation.

Quartogram	f(4)	Trigram	f(3)	Digram	f(2)	Symbol	f(1)
0000	121	000	141	00	170	0	204
0001	20						
0010	19	001	29				
0011	10						
0100	17	010	20	01	34		
0101	3						
0110	9	011	13				
0111	4						
1000	20	100	28	10	33	1	57
1001	8						
1010	1	101	5				
1011	3						
1100	11	110	13	11	23		
1101	2						
1110	4	111	10				
1111	6						

Table 8.3: N-grams frequencies for the binary stress variable

Discussion There aren't many statistical methods that use the n-grams method. One of them is presented in section 8.5. Let's just remind that the main difference between these two methods is that higher order transitional matrices do not consider intermediary states, but that n-grams do. The following statistical techniques employ the former method.

8.2 Higher order dependence using auto-correlations

Determining order with classical statistics is rather straightforward. It was previously shown that a correlation coefficient may be applied on transitional frequency matrices of binary variables and its significance is assessed by the Pearson correlation test (see section 5.1.1). These two statistics are employed to determine the dependence upon previous states.

Methods Higher order auto-correlation coefficients are computed following equation 5.1, computed on higher order transitional frequency matrix. Testing significance of the transitions may be performed using the Pearson correlation coefficient test.

Testing significance of the transitions of configurations may be performed using either the χ^2 or the likelihood ratio chi-square G^2. The p-value will determine if the transitions depend or not on the specified lag k.

Example Figure 8.1 [1] represents the correlation coefficient for the three binary variables stress, emotionality and familiarity of situations. Coefficients are computed for lags 1 to 10. The significance test is also performed; alpha level is set at $\alpha = 0.005$, because there are 10 statistical tests involved. Significant auto-correlations (those that are significantly different from 0) are marked with a diamond symbol.

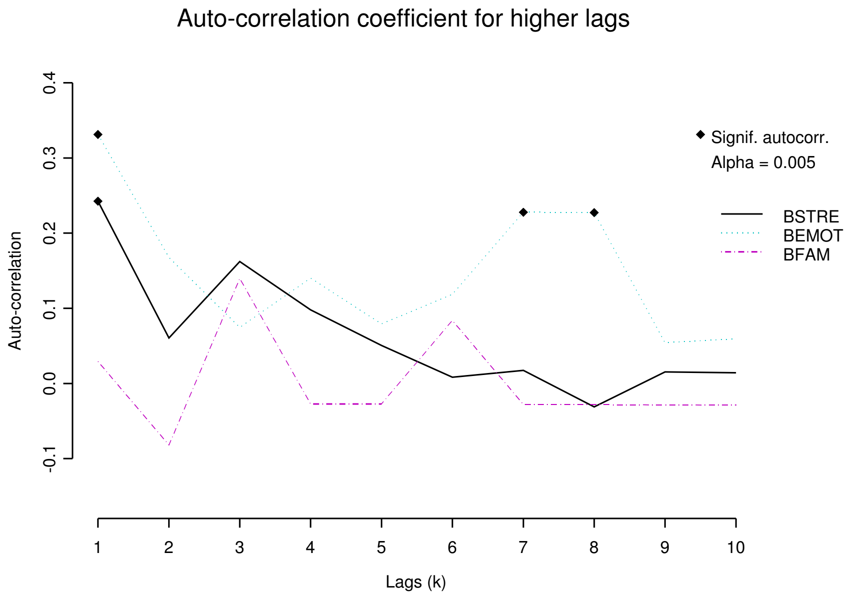

Figure 8.1: Binary autocorrelation coefficients on BSTRE, BEMOT and BFAM

For the stress variable BSTRE, the correlation coefficient is at its highest for $k = 1$, and it is significant. It then drops to a low insignificant level, before rising just below r=0.20 for lag 3 and then slowly decreases towards a 0 coefficient for all other superior lags, which are all insignificant. These results suggest that the stress level of this subject is a process depending on the past state only.

Concerning the emotionality variable BEMOT, more lags seem to be significant: lags 1, 7 and 8. The emotionality level is associated with its immediate past level as well as the 7-the and 8-the preceding ones; this is explained by the frequency of self-reports, set at 7 times a day; it then implies a circadian emotionality rhythm for this subject. The familiarity of situations (BFAM) auto-correlation coefficients never each significance; this result is mainly explained by the presence of a very high frequency of the $BFAM_{0\to0}$ transition at any lags; thus whatever lag is being examined yield a very low auto-correlation coefficient.

Discussion Auto-correlation is an important statistic for determining the order of a sequence of states. For continuous variables it is a fundamental analysis (Kendall & Ord, 1990;

[1] Produced with the S-Plus command:

```
>  BACplot(cbind(BSTRE,BEMOT,BFAM))
```

Lag	χ^2	d.f.	p-value
1	82.09	36	**0.00**
2	59.03	36	**0.01**
3	74.57	36	**0.00**
4	37.34	36	0.41
5	34.86	36	0.52
6	40.70	36	0.27
7	55.35	36	**0.02**
8	39.12	36	0.33
9	42.80	36	0.2
10	37.41	36	0.4

Table 8.4: Testing order of BFAM-BEMOT-BSTRE configuration using chi-square statistic

Gottman, 1981). We showed that for binary variables such coefficient is available. On the one hand there is no such coefficient for purely nominal variable; therefore it could not be applied to configurations. On the other hand, investigators may rely on chi-square statistics to assess the order of their system (next section).

8.3 Higher order dependence using chi-square statistics

The absence of auto-correlation coefficients on categorical data should not prevent researchers to delve the question of higher order Markov chain processes. On nominal variables, including configurations, chi-square statistics was shown to be an appropriate test for testing the dependence between two states. This statistic is thus employed in this section to assess the dependence upon previous states.

Methods Testing significance of the transitions of configurations may be performed on transitional frequency matricesat lag k using either the χ^2 or the likelihood ratio chi-square G^2, as described in section 5.1.3. The associated p-value will determine if the transitions depend or not on the specified lag k.

Example The χ^2-test on transitional matrices for the BFAM, BEMOT and BSTRE configuration is applied for lags 1 to 10. Table 8.4 shows the results (in bold are significance dependences).

Interestingly results show that the three first lags are significant, as well as the seventh lag. It means that configurations are best predicted by using knowledge from the last three states and the seventh the subject was in. The fact that the seventh order emerges is no surprise: observations were taken 7 times a day. Therefore the dynamics of this subject is surely governed by certain daily patterns.

Discussion Even if there is no auto-correlation coefficient for purely nominal variables, investigators may rely on chi-square statistics to assess the order of a dynamical system. We showed an example of such analysis on configurations.

As expressed often in this work, a significant chi-square test never indicate the direction of the relationship between variables. Two significant tests could represent completely opposite patterns of transitions, so researchers are again encouraged to examine the transitional matrices, either in numerical or graphical forms.

8.4 Index of complexity for past states influences

The index of complexity of transitions, presented in section 6.3.1 indicates how transitions from one state to another are ordered (for a small index) or disordered (for a high index). The index of complexity of transitions may also be interpreted as sign of determinism. The index was employed on first-order transitions, by looking the immediate preceding state. But the index can be applied for higher orders. We describe in this section how it may help determine higher order Markov processes.

Method The procedure for exploring higher order influences using index of complexity of transitions is similar to the one using the auto-correlation coefficient: investigator computes transitional matrices for lags 1 to 10 and then applies the index of complexity formula on each of these matrices.

Figure 8.2 shows the results on the deterministic third-order sequence 1,1,0, . . . , previously described.

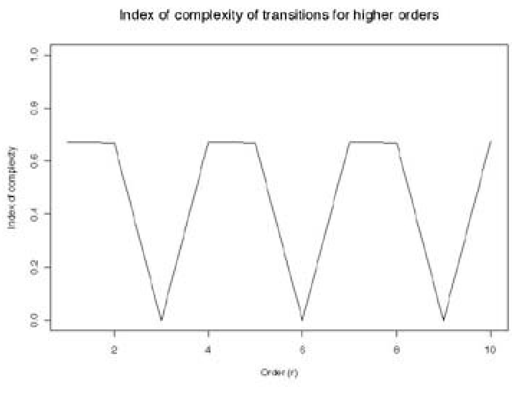

Figure 8.2: Index of complexity for higher orders (3-cycle example)

The figure is interesting because it shows that when considering transitions of the 1st and 2nd order the complexity of the transitions is pretty high. There would not be any order in the sequence of states, if the investigator stopped his search here. But looking one step behind reveal the true structure of the transitions: it is completely deterministic, and shows a cycle of period 3. The figure also shows that there seems to have cycles of period 6 and 9, but these are artifacts given by the presence of the smaller cycle of period 3.

Let's compute now the index of complexity of transitions on greater lags for the dichotomous stress, emotionality, familiarity of situations, and the corresponding configuration. Figure 8.3 summarizes their index of complexity of transitions for higher orders, up to $k = 10$.

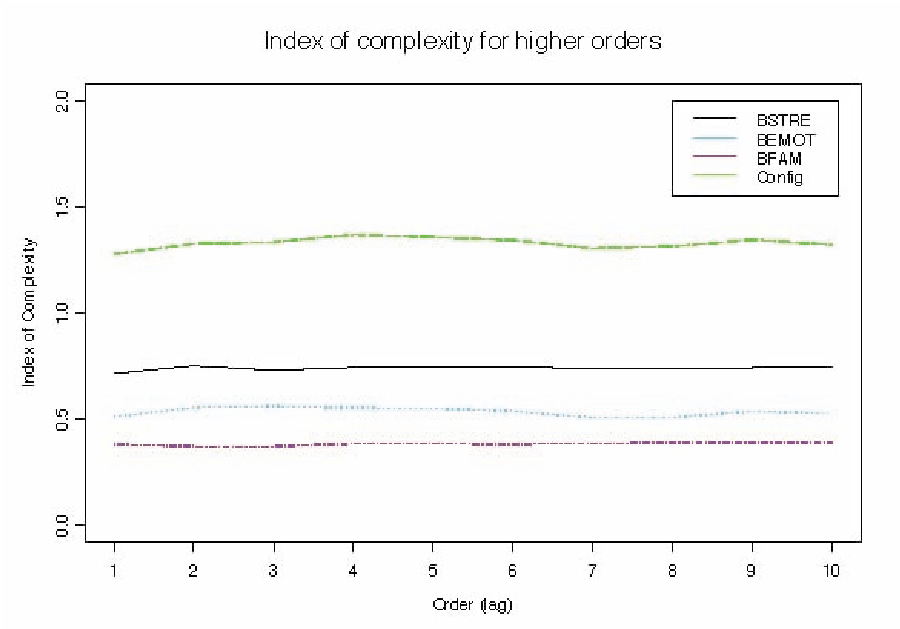

Figure 8.3: Index of complexity for higher orders (BFAM, BEMOT, BSTRE and Config)

The figure clearly indicates that considering higher orders in the sequences does not change the complexity of the system; it remains constant whatever the order. Looking at the past states does not contribute in a reduction of the surprise about the present state.

Discussion The suggested procedure is for exploratory purposes only. It should reveal insight about processes, by determining the complexity of transitional matrices at higher lags. However it suffers from two major limitations.

The first problem relates to its descriptive nature; since it does not statistically determine significance, it can not be used to make inferences. The second limitation concerns its lack of discrimination. This measure seems to work best with more clear-cut processes. The metric of this measure should be investigated (for example using Monte-Carlo simulations).

8.5 Information theoretical methods for higher order

A fruitful method for determining the order in categorical time series data rely on information theoretical methods, and more specifically on entropy and conditional entropy (cf section 6.1). The rationale behind this method is that the information (as computed by the entropy) brought by past state symbols should significant increase if the sequence really depends on past states. The building block for the computations is the n-gram (section 8.1.2, and not the transitional frequency matrix.

Method Attneave (1957) and Gottman and Roy (1990) suggest the following steps for testing the higher-order dependence of a data sequence of categorical nature:

1. Compute the entropy H_i for each i-gram
2. Compute the entropy difference D_i, between the entropy of i-gram (H_i) and entropy of $i-1$-gram (H_{i-1}); that is D_i = H_i - H_{i-1}, which is equivalent to the conditional entropy of order k
3. Compute the difference between these entropy differences: T_i = D_i - D_{i+1}
4. Finally compute χ^2 = $2N \log_e 2T_i$, with $m^{i-1}(m-1)^2$ degrees of freedom

It is understood that it is the difference of the difference of entropies between H_i - H_{i-1} and H_i - H_{+-1} (therefore T_i) that contributes to the increase of information about transitions. The computation becomes clearer when thought in terms of conditional entropy; it is the difference of conditional entropy that is finally tested.

On a computational matter the procedure may be simplified to a more straightforward sequence of steps, because T_i = D_i - D_{i+1}, where D_i = H_i - H_{i-1} and D_{i+1} = H_{i+1} - H_i. Therefore

$$T_i = 2 * H_i - H_{i-1} - H_{i+1} \tag{8.3}$$

by letting H_0 = 1.

The hypotheses that are tested are stated as following (Bavaud, 1998):

$H_0(k)$: the process is a k-th order Markov chain
$H_1(k)$: the process is a $k+1$-th order Markov chain

We reject $H_0(k)$ at the usual α-level with χ^2 = $2N \log_e 2T_i$, with (N is the length of the sequence) and $m^{i-1}(m-1)^2$ degrees of freedom.

Bavaud (Bavaud, 1998) shows that order greater than $\log_2 n$ (n the length of the sequence) are subject to "side-effects" because of the finite length of sequences; any higher order tests should be use with caution. For our example, with $n = 261$, this effect is manifest starting from $\log_2 261 = 5.56$; that is, from the sixth lag and up, the tests become less valid.

Examples

Binary stress variable Let's see how the procedure is applied on our dichotomized stress example (table 8.5).

Lag	H(i)	Delta H(i,i-1)	T(i)	χ^2	d.f.	p-value
1	0.757	0.243	0.958	345.167	1	**0.00**
2	1.472	0.472	-0.001	-0.413	2	1.00
3	2.188	1.188	0.037	13.127	4	**0.01**
4	2.868	1.868	0.048	17.180	8	**0.03**
5	3.499	2.499	0.019	6.676	16	0.98
6	4.112	3.112	0.056	19.939	32	0.95
7	4.668	3.668	0.074	26.003	64	1.00
8	5.151	4.151	0.041	14.370	128	1.00
9	5.593	4.593	0.079	27.794	256	1.00
10	5.956	4.956	0.047	16.489	512	1.00

Table 8.5: Testing k-order using entropy-based method (dichotomous stress variable)

We can observe the increasing amount of entropy and the associated difference of entropy with respect to the last lag, as the size of lag increases. It also shows that the amount of information does not increase significantly.

Results show that there are three significant lags, $k = 1$, $k = 3$ and $k = 4$. It informs that the present stress of the subject is significantly influenced by the stress level one, three and four moments before.

Binary emotionality variable The following table (8.6) shows the testing of the higher order influences for the dichotomous emotionality variable. Results indicate that only the first order is significant, whereas the second order almost reached significance (p-value = 0.06). Contrary to previous analyses this method would point to a first-order Markov chain process only.

Binary familiarity of situations variable Testing the higher order dependence on the familiarity of situations variable indicate a first-order Markov chain process, only the first lag being significant. The result contradicts results from the auto-correlation coefficients analyses for which no lag where significant.

Configurations Unfortunately computing the higher order dependence with this method is practically impossible, because the number of degrees of freedom grows exponentially. For example, testing order $k = 5$ on configuration having 8 possible states gathers $(5-1)^4 * (8-1)^2 = 200704$ degrees of freedom; the number of available data is needless to say unsufficient.

Discussion We have shown in this section another method for testing the order of categorical time series data. It relies on entropy computations and the n-grams. The main advantage

Lag	H(i)	Delta H(i,i-1)	T(i)	χ^2	d.f.	p-value
1	0.579	0.421	0.932	336.020	1	**0.00**
2	1.090	0.090	0.016	5.598	2	0.06
3	1.585	0.585	0.013	4.600	4	0.33
4	2.068	1.068	0.024	8.507	8	0.39
5	2.527	1.527	0.036	12.881	16	0.68
6	2.950	1.950	0.051	17.962	32	0.98
7	3.322	2.322	0.052	18.455	64	1.00
8	3.642	2.642	0.033	11.660	128	1.00
9	3.929	2.929	0.032	11.116	256	1.00
10	4.184	3.184	0.010	3.407	512	1.00

Table 8.6: Testing k-order using entropy-based method (dichotomous emotionality variable)

Lag	H(i)	Delta H(i,i-1)	T(i)	χ^2	d.f.	p-value
1	0.390	0.610	0.988	355.962	1	**0.00**
2	0.768	-0.232	0.007	2.552	2	0.28
3	1.139	0.139	0.009	3.366	4	0.50
4	1.500	0.500	0.006	1.979	8	0.98
5	1.856	0.856	0.000	0.041	16	1.00
6	2.211	1.211	0.004	1.596	32	1.00
7	2.563	1.563	0.003	1.203	64	1.00
8	2.910	1.901	0.005	1.797	128	1.00
9	3.253	2.253	0.009	3.192	256	1.00
10	3.587	2.587	0.005	1.842	512	1.00

Table 8.7: Testing k-order using entropy-based method (dichotomous familiarity of situations variable)

of this approach is that the information theoretic framework is well known and should be by now easily understandable.

The main drawback is there is some serious caution about the "very" high order testing, because the more lags are taken into account the greater the degrees of freedom. Testing configurations order is strictly unthinkable with this method.

8.6 Conclusion

With this chapter closes the set of specific methods for tackling the most important questions of time series analysis. The various methods for assessing the influences of past states on the present state of the system were presented. They rely either on transitional matrices computed at higher lags, or on the n-gram method. Even if auto-correlation coefficients could only be computed on binary variables, other methods based on chi-square statistics or entropy were described. Global higher order Markov chain dependence was presented, as well as specific transitions. Researchers now can proceed to test whatever hypothesis they hold about their dynamical system. But there remains a general methodology that may answer most of the previous questions into a single coherent framework. It is called *loglinear modelling* , and is presented in the next chapter.

Chapter 9

Modelling dynamics

Summary. *Loglinear models allow researchers to answer most of the questions previously described. Using this coherent framework they can tackle questions of higher order Markovian dependence, stationarity of dynamics, multivariate temporal dependence (or "cross-dependence") and the likes. We present in this section the general loglinear methodology, and describe some prototypical examples. Graphical models further allow to represent the network of dependences in a powerfully suggestive compact format.*

"...in the last 20 years or so, the advent of loglinear modelling has revolutionized the multivariate analysis of categorical data" (DeMaris, 1992, p.1)

Loglinear modelling belongs to the very broad class of models called *General Linear Models* . They are an extension of analysis of variance applied to contingency tables. They are said to be linear because the expected values in contingency tables are predicted using linear equations of parameters. But the parameters themselves may be of various sorts, resulting from various transformations: logarithms, logit, probit, and the likes. The choice of models depend on the type of data.

Loglinear models provide a systematic approach to the analysis of complex multidimensional tables and provide estimates of the magnitude of effects. Consequently the relative importance of various effects may be assessed (Everitt, 1992). "This development [of logit model] has brought coherence to the problem by providing a single multivariate model having sufficient flexibility to handle contingency-table, and disaggregated, data, as well as sets of predictors measured at any combination of levels" (DeMaris 1992, p.1)

Cross-tabulations of 3 and more variables having many categories becomes cumbersome. Loglinear models provide an integrated way to deal with these situations.

Loglinear models are mostly used for assessing the interdependencies between variables on a synchronical level. Statisticians have shown that these models may be transposed to contingency tables representing transitions (Bishop et al., 1975; Fienberg, 1980; Gottman & Kumar Roy, 1990).

Loglinear models for categorical time series analysis are used to assess the effects of *contextual designs* (Gottman & Kumar Roy, 1990). Suppose a researcher wants to determine if a particular psychological intervention has any effects on consequent dynamics of psychiatric patients. He would compute the transitional frequency matrices of patients for interventions A and B; loglinear modelling will provide the necessary statistical power to determine if interventions produced any changes or differences in the dynamics as well as the extent of these effects.

In fact, any comparison of transitional frequency matrices can be analyzed using loglinear modelling. It may be used to assess the *stationarity* of dynamics, as in section 7.1. It may assess the *homogeneity* of two dynamics (that is, their similarity or difference), thus revealing if two patients exhibit similar patterns of transitions. It may also be employed for determining the Markovian *order* of the dynamics, as in section 8. And finally it may be used to assess the temporal cross-dependency between groups of variables (later described in section 9.6).

Within loglinear modelling models there are other different types of models, depending on the type of data to be analyzed:

1. logit models
2. linear-logistic
3. polytomous regression

In this work we will focus on general loglinear modelling.

Method The general principle behind loglinear modelling is to construct a *linear* model of the logarithm of the expected cell frequencies in a contingency table. Under the assumption of independence, the expected cell frequencies are given by equation 9.1 (which is the same as equation 5.9):

$$\hat{m}_{ij} = \frac{x_{i+}x_{+j}}{N} \tag{9.1}$$

Taking the logarithm on each side of the equation (and recalling the mathematical property that the logarithm of a product of two numbers is the addition of the logarithm of these two numbers), we get equation 9.2:

$$\log m_{ij} = \log x_{i+} \log x_{+j} - \log N \tag{9.2}$$

which corresponds to a model with a constant $\log N$, an effect from the i-th category of the first variable (on rows) and an effect from the j-th category of the second variable (on columns).

Equation 9.2 may be re-expressed not directly in terms of logarithm of frequencies, but rather in more general fashion, using parameters (cf. equation 9.3):

$$\log m_{ij} = u + u_1 i + u_2 j \tag{9.3}$$

The u-parameter is simply the average of the logarithm of the expected frequency over all i and j categories; it is referred as the *grand mean* , much like in analysis of variance. The parameters are computed using the following equations:

$$u = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \log m_{ij} \tag{9.4}$$

Similarly we get:

$$u + u_1 i = \frac{1}{J}\sum_{j=1}^{J}\log m_{ij} \tag{9.5}$$

$$u + u_2 j = \frac{1}{I}\sum_{i=1}^{I}\log m_{ij} \tag{9.6}$$

These u parameters are uniquely defined and estimated, because they are subject to certain constraints. These constraints are similar to those of analysis of variance; that is, that the sum of parameters for a given variable is equal to 0. That is:

$$\sum_{i=1}^{I} u_1 i = \sum_{j=1}^{J} u_2 j = 0 \tag{9.7}$$

We now include a fourth parameter, $u_{12}ij$, corresponding to the *interaction* between the two variables. It represents the effects of being cross-classified in category i of the first variable and in category j of second variable.

$$\log m_{ij} = u + u_1 i + u_1 i + u_2 j + u_1 2ij \tag{9.8}$$

The later model is of greater interest for sequence analysis because it implies an interaction between states at different moments.

Example As an example, we will use the transitional frequency matrix for our binary stress variable (cf. table 5.3). Suppose the states at time t and $t+1$ are independent, we compute the expected frequencies as in table 9.1.

	$BSTRE(t+1)$		
	0	1	total
$BSTRE(t)$ = 0	159.28	43.72	203
1	44.72	12.28	57
total	204	56	260

Table 9.1: Expected frequencies for binary stress variable (under independence model

We compute the grand mean by taking the logarithm of each frequencies and summing them over: $u = (\log 159.28 + \log 43.72 + \log 44.72 + \log 12.28)/4 = 3.79$

The effect of being in a given category of the row variable, $u_1 i$ (also called the *main effect*), is the difference between the average log expected frequency in row i and the grand mean u. Hence the effect of being in stress level 0 at time t is given by $(\log 159.28 + \log 43.72)/2 - 3.79 = 4.42 - 3.79 = 0.64$. Similarly the effect of being in stress level 1 at time t is given by $(\log 44.72 + \log 12.28)/2 - 3.79 = 3.15 - 3.79 = -0.64$.

We now compute the effect of being in each j column of the second variable. For stress level 0 at time $t+1$, it is given by $(\log 159.28 + \log 44.72)/2 - 3.79 = 4.44 - 3.79 = 0.65$. Similarly the effect of being in stress level 1 at time $t+1$ is given by $(\log 43.72 + \log 12.28)/2 - 3.79 = 3.14 - 3.79 = -0.65$.

The u-parameters correspond to *departures* from what is expected if variables are independent. The greater the u-parameters, the greater the departure from independence. The results are in accordance with the property that u-parameters should sum to 0. And by a little calculus, it would become evident that the average of the row means (and the average of the column means) of expected values is equal to the grand mean; that is, $(4.42 + 3.15)/2 = 3.79$ (and $(4.44 + 3.14)/2 = 3.79$).

We then proceed to compute the association parameters, $u_{12}ij$, those corresponding to the interactions of the two variables. Like the main effects, they also indicate departures but from both the grand mean and the marginal distributions of each variable. If they were completely unrelated, then it would be sufficient to refer to $\log m_{ij}$. But it may not be the case.

The formula for computing the association parameter is in equation 9.9:

$$u_{12}ij = \log m_{ij} - [u + u_1 i + u_2 j] \tag{9.9}$$

Let's compute the particular parameters for each of the 4 interactions:

$$\begin{aligned} u_{00}00 &= \log 159.28 - [3.79 + 0.64 + 0.65] = 0.000 \\ u_{01}01 &= \log\ 43.72 - [3.79 + 0.64 - 0.65] = -0.000 \\ u_{10}10 &= \log\ 44.72 - [3.79 - 0.64 + 0.65] = -0.000 \\ u_{11}11 &= \log\ 12.28 - [3.79 - 0.64 - 0.65] = 0.000 \end{aligned}$$

In fact they all equal in absolute value 5.96e-008, which is not exactly reflected in the later table because of rounding off. These parameter values are very low, thus reflecting an absence of interaction between the two variables.

9.1 Types of models

As in analysis of variance the degrees of freedom have to be computed. The maximum number corresponds to $i \times j$. Then the constraints imposed on the model reduce the degrees of freedom. Three types of model are defined according to the constraints: the saturated model, the independence model and the intermediary non-saturated model.

The saturated model The *saturated model* is the model where all $i \times j$ parameters must be estimated. This leaves no degrees of freedom for testing the goodness of fit of the model. In this model, the G^2 statistic yields 0, with 0 degrees of freedom.

This model always perfectly reproduces the cell frequencies. This is similar to the case of regression analysis using n parameters to fit an equation of n individuals. Perfect prediction is achieved. This model is not interesting *per se* , because the goal of any analysis to build a model as parsimonious as possible (using a minimum number of parameters) that satisfactorily fit a set of data.

For the stress example, one perfectly predict any cell frequency using the parameters. If we were to predict, say m_{01}, the transition from BSTRE=0 to BSTRE=1, we would simple compute: $\log m_{01} = 3.79 + 0.64 - 0.65 = 3.80055$ (when not rounding the answer), yielding $m_{01} = \exp 3.80055 = 44.72$, the exact expected frequency.

Needless to say that this model is of little interest, as a computed statistical result. It is nonetheless useful as a reference model; since it completely accounts for the "variance" of the matrix, subsequent models are to compared with it. The smaller the difference between a specified model and this one, the better the fit (and conversely).

The independence model Models that postulate no interaction between variables are called independence models. The equation of the model is of the form given in equation 9.3. These models estimate only the grand mean parameter (u) and the individual main effects (u_i, u_j, etc).

Intermediary non-saturated models Models that use less parameters than the fully saturated models and more than independence models are usually searched for. We will later see how to select models that provide the best fit of data.

Hierarchical, non-hierarchical and conditional models There are three types of structures regarding loglinear modelling: *hierarchical, non-hierarchical* and *conditional models* . The most usual is the hierarchical type. It implies that if any interaction parameter is included in a model, then all lower-order parameters must be included as well. For example, the presence of parameter $u_{13}ik$ implies that parameters u_1i and u_3k are also included. In non-hierarchical models, this constrain does not exist, so interaction terms may be included without the lower-order terms.

Conditional models (or nested) models (Gottman & Kumar Roy, 1990) are models that create conditional main effects for one variable nested within each level of another explanatory variable. They are based on multiplicative interactive terms (such as in logit models), which can be dropped if necessary, thereby simplifying the interpretative task. The major limitation of this type of models relates to the necessity of specifying a dependent variable, a characteristic similar to causal modelling.

9.2 Estimating expected frequencies

From sample data researchers must proceed to generalize results to the population. Therefore parameters have to estimated. This is mostly performed through the maximum likelihood estimates, which determine values that parameters are most likely to be observed, given the underlying data distribution.

The probability distribution of data are known for any of three sampling plans that are most frequently used: the *multinomial distribution* , the *Poisson distribution* and the *product multinomial distribution* . Multinomial distribution is characteristic of surveys: the total number of cases is fixed in advance and subjects (or observations) are cross-classified according to their joint distribution on the variables. The Poisson distribution results when the temporal interval of observation is fixed, but not the number of observations. The product multinomial distribution is obtained when the number of cases for each category but the dependent variable is fixed.

Fortunately for researchers these three sampling plans yield the same estimates of m_{ij} (Fienberg, 1980).

For two-dimensional contingency tables, the expected frequencies are obtained in a straightforward manner, as in χ^2-statistics. For higher dimensional tables, expected values can not be directly computed. Instead one must use an indirect method such as the Iterative Proportional Fitting or the Newton-Raphson procedures (Fienberg, 1980; Everitt, 1992). Statistical software compute and print them on request for any given model, so we will not go into lengthy details on how expected frequencies are obtained.

9.3 Loglinear models for three-way tables

We now consider three-way contingency tables formed by adding a third variable. The model gets both more complex and more interesting. There are now three main effects, three terms representing associations between pairs of variables and one term the three-way interaction between the three variables. The general equation becomes 9.10:

$$\log m_{ijk} = u + u_1 i + u_2 j + u_3 k + u_{12} ij + u_{13} ik + u_{23} jk + u_{123} ijk \tag{9.10}$$

As before all parameters sum to zero over any subscript i, j or k.

Each set of two-way association parameters represents the conditional association between two variables, controlling for the third. For example, the $u_{12}ij$ parameter gives the association of the first two variables, given every level k of the third variable.

Estimated expected values Adding supplementary variables to the model influences how expected frequencies are estimated. Under a model of independence, these are given as in equation 9.11:

$$\hat{m}_{ijk} = \frac{m_{i++} m_{+j+} m_{++k}}{N^2} \tag{9.11}$$

These are maximum-likelihood estimate for any of the three sampling distribution with one set of one-dimensional marginals totals fixed.

If we consider a model where the first and second variables are independent conditionally upon the third, the MLE of the expected frequencies become (equation 9.12):

$$\hat{m}_{ijk} = \frac{m_{i+k} m_{+jk}}{m_{++k}} \tag{9.12}$$

These are maximum-likelihood estimate for any of the three sampling distribution with one set of one-dimensional marginals totals fixed, as long as the total m_{ij+} is not fixed.

When considering models with no-second interaction, expected values can not be directly computed; instead one must use an indirect method such as the Iterative Proportional Fitting or the Newton-Raphson procedures(Fienberg, 1980; Everitt, 1992) cited previously.

9.4 Goodness of fit and selection of models

Once the expected values have been estimated the goodness of fit of a model can be computed using either the X^2 or G^2 statistics. If the model fitted is correct and the total sample large enough (about ten times the number of cells in the table (Fienberg, 1980)), both statistics have approximate χ^2 distributions with degrees of freedom given by formula 9.13:

$$d.f. = \#cells - \#parametersfitted \tag{9.13}$$

The beauty of the G^2 statistic is that is may not only assess the goodness of fit of a model, but also compare models together. This is possible because this statistic has additive properties. Two hierarchical models, A and B, may be compared by subtracting their G^2, $G^2(A) - G^2(B)$, and their degrees of freedom, d.f.(A) - d.f.(B), thus yielding a new G^2 with given degrees of freedom. This is then tested for significance at an α-level. If the difference is significant (or not), there the two models are said to be different (or not).

This property makes it ideal for selecting the best models that fit data. "As in fitting regression models, we have to balance two contradictory goals or objectives. On the one hand, we would prefer a model complex enough to afford a reasonably good fit of the data. On the other hand, we want a model that is relatively simple to interpret and that is parsimonious - one that smoothes rather than overfits the data" (Gottman & Roy, 1990, p.133).

Four criteria serve as guidelines when selecting models (Gottman & Kumar Roy, 1990):

- *parsimony* : models should contain the fewest number of parameters, yet provide a sufficient goodness of fit;
- *interpretation* : models should be easily and meaningfully interpretable;
- *significant effects* : terms included in the model should be significant;
- *goodness of fit* : models can be selected according to the increase of the goodness of fit;

We remember that the full saturated model completely predicts the cell frequency, and has a $G^2 = 0$ with 0 degrees of freedom. Consequently when assessing goodness of fit, all subsequent models are implicitly compared with the saturated model.

Model building then becomes more apparent. One may either use a "forward" or a "backward" selection approach, like in regression analysis.

In a "forward" selection approach, one starts with the independence model and then progressively adds association parameters that significantly increase the fit of the model. It first starts by testing all possible first-order interactions; it then selects the one that produces the most important and significant increase in fit. If it finds one it is added to the model. The other first-order interactions are again tested, and terms are added if they are significant. The procedure continues with higher-order interactions until no other terms can be added, yielding a final model that best fits the data.

In "backward" selection, one starts from the saturated model. It first tries to remove higher-order interaction terms; if a term is removed if it *does not* yield a significative difference. The procedure is repeated until the removal of a term yield a significative difference; it then indicates that the goodness of fit is too different from the saturated model.

These two approaches may not yield the same model, and no selection method is said to be better than the other.

9.5 Example: modelling stationarity

Let's examine how loglinear modelling helps test whether a sequence of observations is stationary or not. It is illustrated using the binary stress variable of the same subject as before.

The first step consists in building the two transitional frequency matrices one for each period. These were shown in table 7.2.

There are three variables in the tested models: BSTRE(t) (variable [1]), BSTRE(t+1) (variable [2]) and Period (variable [3]). Each variable has two categories: 0 and 1 for stress, 1 and 2 for the period. The null hypothesis to be tested is the following: there is no difference between transitional frequencies during period 1 and 2, but there is a dependence between state at time t and time $t+1$. It corresponds to a model with an interaction between variables 1 and 2, conditionally independent from variables 3. It is formally and compactly written as [3][12] - or [12][3] the order of the variable does not have any importance; since we are dealing with hierarchical models, the presence of an interaction term implies the presence of lower order terms (here such as [1], [2]), so the complete specified model includes in fact terms [1], [2], [3], [12].

Taking advantage of computer speed and power, all possible models are computed, as shown in table 9.2. The complete results allow to compare models afterwards.

Model	$\mathbf{X^2}$	$\mathbf{G^2}$	**df**	**p**
[1][2][3]	18.25	15.89	4	0
[1][23]	15.92	14.43	3	0
[2][13]	16.29	14.79	3	0
[3][12]	2.07	2.08	3	0.56
[12][13]	0.97	0.97	2	0.62
[12][23]	0.61	0.61	2	0.74
[13][23]	14.70	13.33	2	0
[12][13][23]	0	0	1	0.97

Table 9.2: Loglinear models for two periods of binary stress (subject: FA7)

Results indicate that the model[3][12] produces a X^2 test of 2.07, $G^2 = 2.08$, has 3 degrees of freedom, yielding a p-value of 0.56. The test of independence is not significant, so we must accept the hypothesis that the two transitional frequency matrices of the two periods are similar, while the transitions between the two lags $t, t+1$ are significant. Therefore the dynamics is considered as stationary.

Question now arises if this is the best model given the data. May be a model including other interactions would provide a best fit of data. Looking at table 9.2 one may be tempted to choose a model showing a significant p-value. But let's recall that these models are implicitly compared with the full saturated model, that provides the best - and exact - fit to the data: it has a $X^2 = G^2 = 0$, with 0 degree of freedom. So we are looking for a model that is non-significant, since a significant model implies a departure from the best model.

Three other models show non-significant p-values: [12][13], [12][23] and [12][13][23]. They all include the [12] term, which confirms the transitional state dependence $t, t+1$. The other terms [13] and [23] are "artifacts" of our transitional frequency matrices, because there are as many BSTRE(t) and BSTRE(t) in period 1 and 2. They should be included as constraints in the models.

The same loglinear models are then tested using the a forward model selection procedure.

The stationarity of the binary stress variable is now tested using the forward selection procedure. Here is the output given from the MIMWIN 3.0 Graphical Model software by David Edwards (Edwards, 1995).

```
MIM->fact a2b2c2
MIM->labels a "BSTRE(t)"  b "BSTRE(t+1)"  c "Period"
MIM->statread abc
DATA->90  80  15  18  16  18  9  14  !
Reading completed.
MIM->model a,b,c
MIM->stepwise f
Deviance:      15.8939 DF: 4
Non-coherent Forward Selection
Decomposable models, Chi-squared tests.
DFs adjusted for sparsity.
Critical value:   0.0500
Initial model: A,B,C
Model: A,B,C
Deviance:  15.8939 DF:   4 P:  0.0032
    Edge        Test
   Added   Statistic DF          P
    [AB]     13.8188  1      0.0002 +
    [AC]      1.1033  1      0.2936
    [BC]      1.4606  1      0.2268
Added edge [AB]
Model: C,AB
Deviance:   2.0751 DF:   3 P:  0.5570
    Edge        Test
   Added   Statistic DF          P
    [AC]      1.1033  1      0.2936
    [BC]      1.4606  1      0.2268
No change.
Deviance:       2.0751 DF: 3
Selected model: C,AB
```

Letter A stands for BSTRE(t), B for BSTRE(t+1) and C for the period. It firsts starts by considering the independence model [A][B][C]; this model has a deviance (G^2) of 15.8939, with 4 degrees of freedom, which is highly significant (p=0.0032). First-order interactions, [AB][AC][BC], are then tested; only the [AB]-term is significant (the comparison between the [A][B][C] and [A][B][C][AB] models yields a deviance of $G^2 = 13.8188$, $df = 1$, which is significant with $p = 0.0002$), so [AB] is added to the model. The two other first-order

1st period

	$BEMOT(t+1)$	
$BEMOT(t)$	0	1
0	88	14
1	15	13

2nd period

	$BEMOT(t+1)$	
$BEMOT(t)$	0	1
0	116	6
1	6	2

Table 9.3: Transitional frequencies for two periods (BEMOT)

interaction terms are tested, but none are significant. Of course the second-order term [ABC] is not tested, because it corresponds to the full saturated model which provides the exact fit of data.

The selected model is graphically represented in figure 9.1.

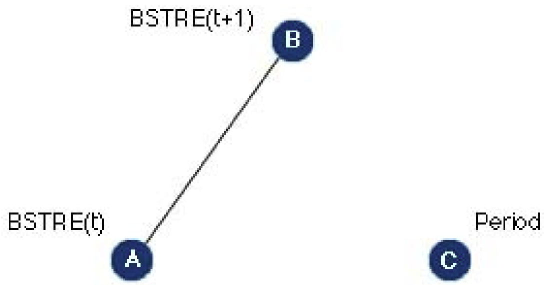

Figure 9.1: Graphical model of stationarity model of binary stress

We now examine another example, for the dichotomous emotionality variable (same subject). Her transitional frequency matrices are displayed in table 9.3.

We first examine all models produced by the combination of the three variables. Table 9.4 shows all necessary statistics. Letter A and number [1] stand for BEMOT(t), B and [2] for BEMOT(t+1) while C and [3] for the Period. Results indicate that only one model is statistically non-significant with respect to the full saturated model: [12][13][23] ($G^2 = 0.03$, $df = 1, p = 0.87$). We conclude that the dynamics is non-stationary, since there are significant interactions between the three variables.

Results are clear enough, we further analyze the stationarity of the transitions using the forward selection procedure. Results are exactly the same as previously shown.

```
MIM->fact a2b2c2
MIM->labels a "BEMOT(t)"  b "BEMOT(t+1)"  c "Period"
MIM->statread abc
```

Model	$\mathbf{X^2}$	$\mathbf{G^2}$	**df**	**p**
[1][2][3]	62.27	41.88	4	0
[1][23]	35.66	29.38	3	0
[2][13]	34.69	28.33	3	0
[3][12]	18.97	20.15	3	0
[12][13]	6.48	6.59	2	0.04
[12][23]	7.49	7.65	2	0.02
[13][23]	19.52	15.83	2	0
[12][13][23]	0.03	0.03	1	0.87

Table 9.4: Loglinear models for two periods of binary emotionality (subject: FA7)

```
DATA->88  116  14  6  15  6  13  2  !
Reading completed.
MIM->model a,b,c
MIM->stepwise f
Deviance:      41.8835 DF: 4
Non-coherent Forward Selection
Decomposable models, Chi-squared tests.
DFs adjusted for sparsity.
Critical value:   0.0500
Initial model: A,B,C
Model: A,B,C
Deviance:  41.8835 DF:   4 P:  0.0000
    Edge         Test
   Added   Statistic DF          P
    [AB]      21.7378  1      0.0000 +
    [AC]      13.5558  1      0.0002 +
    [BC]      12.4986  1      0.0004 +
Added edge [AB]
Model: C,AB
Deviance:  20.1457 DF:   3 P:  0.0002
    Edge         Test
   Added   Statistic DF          P
    [AC]      13.5558  1      0.0002 +
    [BC]      12.4986  1      0.0004 +
Added edge [AC]
Model: AC,AB
Deviance:   6.5898 DF:   2 P:  0.0371
    Edge         Test
   Added   Statistic DF          P
    [BC]       6.5898  2      0.0371 +
```

```
Added edge [BC]
Deviance:        0.0000 DF: 0
Selected model: ABC
```

The corresponding model is graphically represented in figure 9.2.

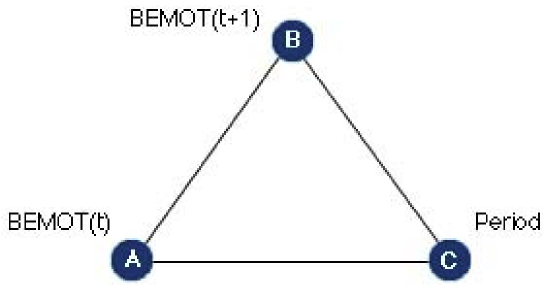

Figure 9.2: Graphical model of stationarity model of binary emotionality

We could easily extend the previous reasoning to transitional matrices of data other than binary; following our configural approach it would be necessary to test the stationarity using the loglinear modelling framework. Unfortunately, for most of our subjects, there are too many zeros in the transitional matrices. Researchers have to be aware that the minimal number of observations required for such analysis is pretty high, because of the computation of two transitional matrices of size $m \times m$; multiply the number of cells times 5 (minimal cell frequency for valid chi-square tests) times 2, which yields $10m^2$. For configurations made up of 3 binary variables ($m = 2^3 = 8$), it would then require 640 observations to perform such analyses. Some better methods for dealing with sparse tables are truly necessary.

9.6 Example: modelling cross-dependence

Modelling order or stationarity are good examples of the power and versatility of the loglinear modelling approach. We now use this technique to model the cross-dependence of variables.

We wonder if the state of many variables at a given moment t influences the state of a variable at time $t+1$. In our psychosocial example, we would like to know if the emotionality, stress or familiarity of situations at time $t + 1$ is determined by the state of these variables at time t.

Method We will perform three loglinear modelling analyses to model how variables emotionality, stress and familiarity of situations are temporally related. The first analysis concerns the dependence between variables BFAM(t), BEMOT(t), BSTRE(t) with BSTRE(t+1), the second analysis concerns these three variables with BEMOT(t+1), while the third analysis relates the same three variables with BFAM(t+1).

The analysis are performed using the Edwards' graphical modelling software (Edwards, 1995), with a forward selection scheme. A graphical model for each analysis is presented, while a final unique figure will synthesize the individual analyses.

Modelling stress cross-dependence Here are the results for the analysis of the model relating BFAM(t), BEMOT(t), BSTRE(t) with BSTRE(t+1), in graphical form (figure 9.3) and numerical output.

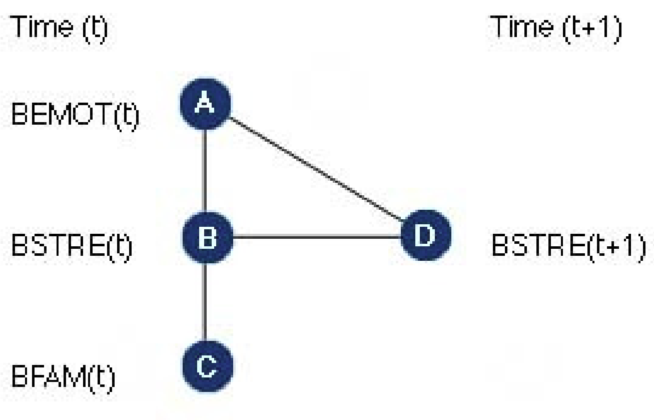

Figure 9.3: Graphical model of predicting BSTRE from BFAM, BEMOT and BSTRE

```
MIM->fact a2b2c2d2
MIM->labels a "BEMOT(t)"  b "BSTRE(t)"  c "BFAM(t)" d "BSTRE(t+1)"
MIM->statread abcd
DATA->5  1  162  28  2  3  15  8  0  0  3  4  8  1  9  11  !
Reading completed.
MIM->model a,b,c,d
MIM->stepwise f
Deviance:     121.2075 DF: 11
Non-coherent Forward Selection
Decomposable models, Chi-squared tests.
DFs adjusted for sparsity.
Critical value:   0.0500
Initial model: A,B,C,D
Model: A,B,C,D
Deviance: 121.2075 DF:  11 P:  0.0000
    Edge        Test
```

```
   Added   Statistic DF         P
    [AB]     69.2245  1      0.0000 +
    [AC]     12.7771  1      0.0004 +
    [AD]     11.2502  1      0.0008 +
    [BC]     23.3892  1      0.0000 +
    [BD]     13.8188  1      0.0002 +
    [CD]      0.1486  1      0.6999
Added edge [AB]
Model: D,C,AB
Deviance:  51.9830 DF:  10 P:  0.0000
    Edge        Test
   Added   Statistic DF         P
    [AC]     12.7771  1      0.0004 +
    [AD]     11.2502  1      0.0008 +
    [BC]     23.3892  1      0.0000 +
    [BD]     13.8188  1      0.0002 +
    [CD]      0.1486  1      0.6999
Added edge [BC]
Model: D,BC,AB
Deviance:  28.5938 DF:   9 P:  0.0008
    Edge        Test
   Added   Statistic DF         P
    [AC]      1.7784  2      0.4110
    [AD]     11.2502  1      0.0008 +
    [BD]     13.8188  1      0.0002 +
    [CD]      0.1486  1      0.6999
Added edge [BD]
Model: BD,BC,AB
Deviance:  14.7749 DF:   8 P:  0.0637
    Edge        Test
   Added   Statistic DF         P
    [AC]      1.7784  2      0.4110
    [AD]      6.3780  2      0.0412 +
    [CD]      1.1039  2      0.5758
Added edge [AD]
Model: BC,ABD
Deviance:   8.3969 DF:   6 P:  0.2104
    Edge        Test
   Added   Statistic DF         P
    [AC]      1.7784  2      0.4110
    [CD]      1.1039  2      0.5758
No change.
Deviance:       8.3969 DF: 6
Selected model: BC,ABD
```

The model that best fit data is [BC][ABD], which integrates two important features. The first feature is synchronical: at time t, emotionality (BEMOT) and familiarity of situations (BFAM) are dependent upon stress (BSTRE), while the two former variables being independent from each other. The second feature is dynamical: stress at time $t + 1$ depends upon both emotionality and stress (but not on familiarity of situations). We then conclude that stress at a given moment depends only on internal variables such stress and emotionality, but not on external factors.

Modelling emotionality cross-dependence Now the results for the analysis of the model relating BFAM(t), BEMOT(t), BSTRE(t) with BEMOT(t+1), in graphical form (figure 9.4) and numerical output.

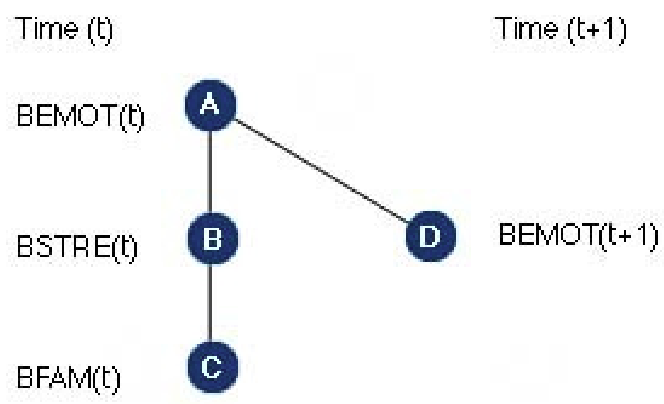

Figure 9.4: Graphical model of predicting BEMOT from BFAM, BEMOT and BSTRE

```
MIM->fact a2b2c2d2
MIM->labels a "BEMOT(t)"  b "BSTRE(t)"  c "BFAM(t)" d "BEMOT(t+1)"
MIM->statread abcd
DATA->5  1  174  16  4  1  21  2  0  0  3  4  8  1  10  10  !
Reading completed.
MIM->model a,b,c,d
MIM->stepwise f
Deviance:     122.4654 DF: 11
Non-coherent Forward Selection
Decomposable models, Chi-squared tests.
DFs adjusted for sparsity.
Critical value:   0.0500
Initial model: A,B,C,D
Model: A,B,C,D
```

```
Deviance: 122.4654 DF:  11 P:  0.0000
    Edge        Test
   Added   Statistic DF          P
    [AB]     69.2245  1      0.0000 +
    [AC]     12.7771  1      0.0004 +
    [AD]     21.7378  1      0.0000 +
    [BC]     23.3892  1      0.0000 +
    [BD]      6.8506  1      0.0089 +
    [CD]      0.0428  1      0.8361
Added edge [AB]
Model: D,C,AB
Deviance:  53.2409 DF:  10 P:  0.0000
    Edge        Test
   Added   Statistic DF          P
    [AC]     12.7771  1      0.0004 +
    [AD]     21.7378  1      0.0000 +
    [BC]     23.3892  1      0.0000 +
    [BD]      6.8506  1      0.0089 +
    [CD]      0.0428  1      0.8361
Added edge [BC]
Model: D,BC,AB
Deviance:  29.8517 DF:   9 P:  0.0005
    Edge        Test
   Added   Statistic DF          P
    [AC]      1.7784  2      0.4110
    [AD]     21.7378  1      0.0000 +
    [BD]      6.8506  1      0.0089 +
    [CD]      0.0428  1      0.8361
Added edge [AD]
Model: BC,AD,AB
Deviance:   8.1139 DF:   8 P:  0.4224
    Edge        Test
   Added   Statistic DF          P
    [AC]      1.7784  2      0.4110
    [BD]      0.9647  2      0.6173
No change.
Deviance:       8.1139 DF: 8
Selected model: BC,AD,AB
```

The model that best fits data is [BC][AD][AB]; as before on a synchronical level (at time *t*) emotionality (BEMOT) and familiarity of situations (BFAM) are dependent upon stress (BSTRE), while being independent from each other. On a dynamical level, emotionality at a given moment, BEMOT(t+1), is determined by the emotionality level the previous moment BEMOT(t). The result implies a "catalytic" process, where the emotionally experienced 2 hours before significantly determines the emotionality the present moment.

Modelling familiarity of situations cross-dependence Now the result for the analysis of the model relating BFAM(t), BEMOT(t), BSTRE(t) with BFAM(t+1), in graphical form (figure 9.5) and numerical output.

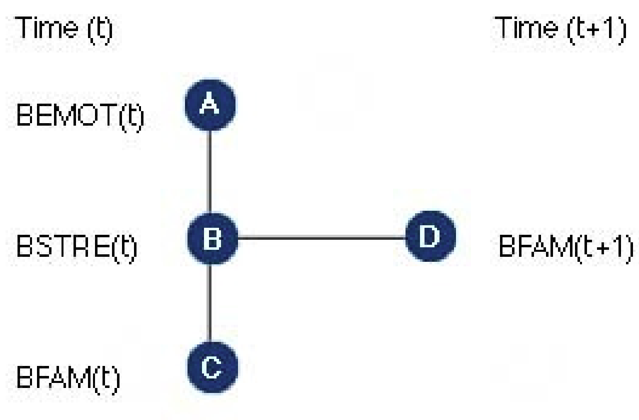

Figure 9.5: Graphical model of predicting BFAM from BFAM, BEMOT and BSTRE

```
MIM->fact a2b2c2d2
MIM->labels a "BEMOT(t)"  b "BSTRE(t)"  c "BFAM(t)" d "BFAM(t+1)"
MIM->statread abcd
DATA->0  6  10  180  0  5  4  19  0  0  1  6  2  7  2  18  !
Reading completed.
MIM->model a,b,c,d
MIM->stepwise f
Deviance:     102.4994 DF: 11
Non-coherent Forward Selection
Decomposable models, Chi-squared tests.
DFs adjusted for sparsity.
Critical value:   0.0500
Initial model: A,B,C,D
Model: A,B,C,D
Deviance: 102.4994 DF:  11 P:  0.0000
    Edge         Test
   Added   Statistic DF          P
    [AB]     69.2245  1      0.0000 +
    [AC]     12.7771  1      0.0004 +
    [AD]      2.2432  1      0.1342
    [BC]     23.3892  1      0.0000 +
    [BD]      4.2255  1      0.0398 +
```

```
    [CD]      0.2115  1      0.6456
Added edge [AB]
Model: D,C,AB
Deviance:  33.2749 DF:  10 P:  0.0002
    Edge        Test
   Added   Statistic DF          P
    [AC]     12.7771  1      0.0004 +
    [AD]      2.2432  1      0.1342
    [BC]     23.3892  1      0.0000 +
    [BD]      4.2255  1      0.0398 +
    [CD]      0.2115  1      0.6456
Added edge [BC]
Model: D,BC,AB
Deviance:   9.8857 DF:   9 P:  0.3598
    Edge        Test
   Added   Statistic DF          P
    [AC]      1.7784  2      0.4110
    [AD]      2.2432  1      0.1342
    [BD]      4.2255  1      0.0398 +
    [CD]      0.2115  1      0.6456
Added edge [BD]
Model: BD,BC,AB
Deviance:   5.6602 DF:   8 P:  0.6852
    Edge        Test
   Added   Statistic DF          P
    [AC]      1.7784  2      0.4110
    [AD]      0.7995  2      0.6705
    [CD]      0.6799  2      0.7118
No change.
Deviance:       5.6602 DF: 8
Selected model: BD,BC,AB
```

The best selected model is [BD][BC][AB], which states once again that on a synchronical level (at time t) emotionality (BEMOT) and familiarity of situations (BFAM) are dependent upon stress (BSTRE), while being independent from each other. On a dynamical level familiarity of situations at a time $t+1$ depends upon the stress level the previous moment t. This may come as a surprise: this external variable is determined by an internal variables (stress). Consequently what this subject experienced at present moment influences two hours later the situations she will be in (or at least the interpretation she will make of it).

Integrating results All three analyses are integrated into a single coherent picture. Figure 9.6 represents the emerging graphical model.

Examination of the graphical model shows the pivotal importance of the internal variables - especially stress -, on both synchronical and dynamical levels. At a given moment t, stress is dependent upon emotionality, and familiarity of situations (while the two other vari-

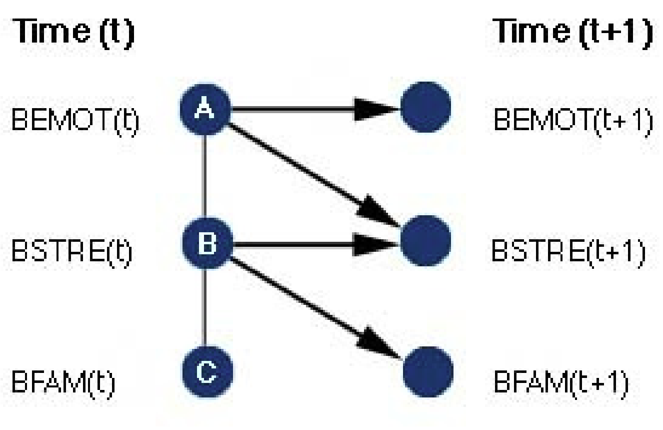

Figure 9.6: Graphical model of cross-predicting BFAM, BEMOT and BSTRE

ables are independent from each other). On a dynamical level, stress at time t significantly determines at time $t+1$ familiarity of situations and itself, while emotionality significantly determines at time $t+1$ stress and itself. Consequently there are two auto-catalytic processes for this subject.

9.7 Discussion

Log-linear modelling is a powerful statistical technique that allow to model cross-dependence between variables at different moments. We have shown how our three variables interrelated between time t and $t+1$.

As previously stated these results, even if coherent as a whole, should be considered with care, as there are many contingency cells that are less frequent than 5.

As emphasized through out this work investigators should always relate results of their analyses with transitional matrices and state transition diagrams such as Karnaugh maps. A significant dependence between two variables never indicates the direction of this dependence. Here we have not expressly done it, since they were amply discussed through out.

9.8 Conclusion

Loglinear modelling provides an integrated framework. We used them to examine how configural variables are co-determined. Combined with a graphical representation, the network of influences can be easily interpreted. Investigators should use this technique combined to Karnaugh maps and examination of transitional matrices to pin point the direction of influences and consequently the interpretation.

Chapter 10

Conclusion

10.1 Summary

We have presented in this work the major methods for analyzing categorical time series data. The effort was undertaken because there was a tremendous lack of resources, especially in software. We programmed many statistical routines and made available to the scientific community.

Not only did existing statistics were programmed but *new procedures* were also developed. Graphical representations of dynamics, such as time series plots and phase spaces, could not apply directly on categorical time series data. We thus invented two new graphics, the *evolustrip* and *Karnaugh map* . The latter was already an established technique in computer science, but we diverted its regular usage and transformed it into a technique for representing the dynamics of categorical data. It functions as a multivariate state transition diagram. The advantages of these graphics, especially the micro-macro perspective brought by the Karnaugh map, were stressed out.

Through out this work we mainly emphasized a *configural* approach. We believe that much processes in psychology are of a multivariate nature. Bio-psycho-social variables are intensely intertwined, each influencing the other. The dense network of interdependence makes it often difficult to distinguish a precise causal path between the variables. The use of configurations abolish such distinction, yield new entities exhibiting emerging properties not found at an univariate level.

We showed how to build configurations, expliciting how to select variables, dichotomize and assemble them. Dichotomization is best achieved at mid-scale. An *a priori* interesting choice was the median, but it artificially increases the entropy of configurations. An measure relying on information theory was suggested to compute the resulting loss of information.

Karnaugh map graphics are the privileged method for representing the synchronical and dynamical properties of a system. Indeed both the structure and the dynamics can be shown using this method. By displaying frequencies of configurations researchers can easily spot the most and least frequent configurations; following the Configural Frequency Analysis approach they represent *types* and *antitypes* .

The use of configurations brings a double focus on *relationships between variables* , the primary concern of χ^2 statistics and loglinear modelling, to relationships between individual cells (or states). This is partly due to the topological properties of Karnaugh maps that further show the *micro- macro-relationships* between variables. Researchers can analyzed the graphics variable by variable (by examining the graphic in lines or columns), or by configurations.

Configurations appear to be the most important multivariate approach when dealing with categorical variables. Its main advantage lies in the integration of many categorical variables into a single entity. Standard analyses such as chi-square statistics may be employed. The main disadvantages relate to the fact that configurations are then treated as a *nominal* variable. The number of categories is then increased manifold, often leading to sparse transitional frequency matrices.

Moreover if care is not taken to ensure a comparison between categories the approach losses all its benefits. For example Karnaugh map is a valuable technique because all configurations are displayed and changes between configurations are related to changes in individual variables. An χ^2 test on a transitional matrix is interesting only if individual configurations are examined to determine the locus of (non-) significance. We emphasized it throughout this work because contrary to analyses on continuous variables, where a statistic always indicates the direction of influences (e.g. a positive or negative correlation), statistics on nominal variables does not indicate this fundamental characteristics.

After having shown suitable graphical representation of time series evolution we then proceeded to the main statistical methods for analyzing the dynamics. We described how to test the *first-order* Markovian dynamics in a sequence of observations, mainly using X^2 and G^2 statistics; we followed by showing how to test specific transitions. Higher-order transitions were later explained and detailed.

We also showed how to examine the *stationarity* of dynamics. It is a methodological requirement that must be met before any statistical analysis; it seems at first a paradoxical statement to require a dynamics not to change for later determining if there is some changes occurring in a system. But it corresponds to a need of stability in the dynamics if statistical tests are to be reliable.

Beside the purely methodological side of the analysis, the search for stability/instability in the dynamics may become an object of inquiry by itself. Researchers may be interested in determining if a dynamics was stable across time, or if it changed during certain period. This was labeled as the search for *phases* . An iterative procedure based on χ^2 statistic was described; it separates the sequence of observations at all possible (reasonable) points and assess the significance of the comparison of dynamics. Plotting the evolution of this statistic as a function of time allows to locate moments of *bifurcation* .

Phases are further explored by plotting results from a technique similar to the moving-average performed on continuous variables: a *moving-mode* , *moving-entropy* and *moving-index-of-complexity* . We suggested to plot both mode and entropy (or complexity), for configurations and individual variables. Researchers then visualize the evolution of the most important state or configuration in segments of definite length (e.g. a day or a week). Periods of stability and instability are then immediately visible.

Loglinear modelling was presented as a general technique for investigating various problems previously treated, such as stationarity, cross-dependence and comparison of dynamics. Models can be visualized using graphical models. The principles and methods for building these models were explained.

10.2 Perspectives

The groundwork foundations for categorical time series analyses have been laid. Researchers should now possess the necessary references and computerized statistical tools to investigate about any theoretical questions.

Despite the scope of this work some questions were left unanswered. These are: a) we did not find (yet) adequate techniques for testing the presence of *periodicity* in categorical time series; B) the *ordinal* type of data was not exploited at all; C) the *forecasting* and *control* perspectives for which clinical psychologists by interested in was not examined. For these questions further statistical development is necessary.

We emphasized a configural approach for the analysis of categorical time series data, along with methods based on information theory. Some fruitful perspectives are foreseen as one would integrate other closely related approaches; we mainly think about cellular automata (Wolfram, 1994) and boolean networks (Kauffman, 1993). Their approach models the evolution of systems using intertwined networks of binary elements (the automata) connected through simple boolean equations. They showed how systems self-organize on a macro-level using simple behavior at a micro-level.

This line of investigation is not unrelated to the work of Thomas in Kinetic logic (Thomas, 1979; Thomas & D'Ary, 1990). We have begun to explore this promising avenue, but time not being compressible at will, the few knowledge we have got is not yet sufficient to allow a full discussion on the matter. But it is THE logical suite of our work.

Finally we could not pass over in silence the increasing prominence of the Internet and the World Wide Web. We believe that many efforts will be directed towards the use of online evaluative instruments. It is almost now a trivial task to set up such instruments. And providing an *immediate feedback* just requires a little more programming knowledge.

Consequently a very challenging task awaits dynamically-oriented psychologists: build interactive systems that allows repeated self-observation of subjects (an Experience Sampling Method) using pocket devices and provide immediate feedback of particular results.

For this to become a reality in the near future important efforts must be engaged towards the development of *robust* statistical procedures. The bootstrap method and its time series counterpart, the *surrogate* method, is a general technique that allows to make valid inferences about some statistics. Applied to time series data, it can test if there is really a k-th order dynamics, or if the underlying distribution accounts for much of the structure of transitions. Bootstrap procedures could be used with index of complexity and χ^2 statistic. It could also address the question of phases: we wonder what is the minimal number of observations required to produce the dynamics we observed. Intensive re-sampling of segment of increasing length would show the sufficient number of observations. This line of thinking shows how researchers can investigate the much debated question of the minimum number of observations.

The ground work has been done *and* the perspectives are stimulating. We hope that the description of statistical analyses and their accompanying statistical routines provided in this work could help dynamically-oriented psychologists find meaningful answers to their problems.

Appendix A

ESM Empirical data

For the sake of coherence and clarity, we deal through out this work with a specific set of empirical data, mostly from a unique subject. It serves as a reference for all analyses to be performed, thus providing a large-scale case study. An Experience Sampling Method research (Lemay, Vuadens, Dauwalder, & Pomini, 1994) was undertaken during spring 1993. The general objective of the research was to understand the bio-psycho-social dynamics of tolerance behavior. We wanted to better understand how various internal and external variables were intertwined with contextual factors such as place and time of interactions to give rise to tolerance behavior.

The instrument was mainly composed of nine questions, seven of them rated on a six-point Likert scale; the seven ordinal variables were:

1. mood (MOOD: negative/positive)
2. emotionality level (EMOT: low/high)
3. stress level (STRE: low/high)
4. pressures and demands from environments (PRESS: low/high)
5. material facilitation (MATF: obstacle/helping)
6. social support (SOCS: negative/positive)
7. familiarity of situations (FAM: unfamiliar/familiar)

The other two nominal variables are the place of interactions (PLACE: at home, at a friend's place, at work, in a public place, other place) and social setting (PEOPLE: alone, partner, friend(s), colleague(s), unknown(s), child(children), other people).

These variables were self-rated seven times a day (every two hours starting from 10:00 am) for a period of 37 days (a little more than 5 weeks), yielding a total of 261 observations. Eight subjects participated in the study: they were all swiss, female students from the University of Lausanne, aged between 21 and 23 years old; they were all singles; four were catholic and four were protestant.

References

Abraham, R. (1990). *Complex dynamical systems. Selected papers by Ralph Abraham, The science frontier express series.* Santa Cruz: Aerial Press.

Allison, P. D., & Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: a comment on gottman. *Psychological bulletin*, *91*(2), 393-403.

Aula, P. S. (1996). Chaos and the double function of communication. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 191-206). Singapore: World Scientific.

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction. An introduction to sequential analysis.* Cambridge: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with sdis and gseq.* New York: Cambridge University Press.

Barton, S. (1994). Chaos, self-organization, and psychology. *American Psychologist*, *49*(1), 5-14.

Bavaud, F. (1998). *Modèles et données. Une introduction à la statistique uni-, bi- et trivariée.* Paris: L'Harmattan.

Bergeret, J. (1991). Ordre et désordre mental. *La Recherche*, *232*(Mai), xx.

Bersier, M. (1997). *Modélisation du changement dans les processus de consultation.* Unpublished doctoral dissertation, University of Lausanne.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis. Theory and practice.* Cambridge: MIT Press.

Bossel, H. (1994). *Modelling and simulation.* Wellesley, Mass: A.K Peters.

Box, G. E., & Jenkins, G. M. (1976). *Time-series analysis. Forecasting and control.* San Francisco: Holden Day.

Brandsttter, H. (1991). Emotions in everyday life situations. Time sampling of subjective experience. In *Subjective well-being. An interdisciplinary perspective* (p. 173-192). Oxford: Pergamon Press.

Castellan, N. J. J. (1979). The analysis of behavior sequences. In R. B. Cairns (Ed.), *The analysis of social interactions. Methods, issues and illustrations* (p. 81-116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, M., & Isen, A. (1982). Toward understanding the relationship between feeling states and social behavior. In A. Hastorf & A. Isen (Eds.), *Cognitive social psychology* (p. 73-108). New York: Elsevier North-Holland.

Combs, A. (1996). Coda consciousness: chaotic and strangely attractive. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 401-411). Singapore: World Scientific.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory.* New York: John Wiley.

Dantzer, R. (1994). *Les émotions.* Paris: Presses Universitaires de France.

Delahaye, J.-P. (1994). *Information, complexité et hasard.* Paris: Hermès.

Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life. The experience sampling method.* Maastricht: Universitaire Pers Maastricht.

Deschamps, J.-C., & Clémence, A. (1987). *L'explication quotidienne.* Cousset: DelVal.

Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (Eds.). (1994). *Analysis of longitudinal data.* Oxford: Clarendon Press.

Edwards, D. (1995). *Introduction to graphical modelling.* Berlin: Springer.

Everitt, B. S. (1992). *The analysis of contingency tables (second edition).* London: Chapman and Hall.

Eye, A. von. (1990). *Introduction to configural frequency analysis. The search for types and antitypes in cross-classifications.* Cambridge: Cambridge University Press.

Eye, A. von, Spiel, C., & Wood, P. K. (1996). Configural frequency analysis in applied psychological research. *Applied psychology*, *45*(4), 301-352.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data.* Cambridge: MIT Press.

Franklin, R. D., Gorman, B. S., Beasley, T. M., & Allison, D. B. (1996). Graphical display and visual analysis. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (p. 119-158). Mahwah, NJ: Lawrence Erlbaum Associates.

Freeman, W. (1991). The physiology of perception. *Scientific American*, *264*(February), 78-85.

Gleick, J. (1987). *Chaos: The making of a new science.* New York: Viking Penguin.

Goldberger, A., Rigney, D., & West, B. (1990). Chaos et fractals en physiologie humaine. *Pour la Science*, *150*, xx.

Golden, L. L., Brockett, P. L., & Zimmer, M. R. (1990). An information theoritic approach for identifying shared information and asymmetric relationships among variables. *Multivariate behavioral research*, *25*(4), 479-502.

Goldstein, J. (1996). Causality and emergence in chaos and complexity theories. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 161-190). Singapore: World Scientific.

Gottman, J. M. (1981). *Time-series analysis. A comprehensive introduction for social scientists.* Cambridge: Cambridge University Press.

Gottman, J. M., & Kumar Roy, A. (1990). *Sequential analysis. A guide for behavioral researchers.* Cambridge: Cambridge University Press.

Gottman, J. M., & Notarius, C. (1978). Sequential analysis of observational data using markov chains. In T. R. Kratchowill (Ed.), *Single subject research. strategies for evaluating change* (p. 287-311). New York: Academic Press.

Haken, H. (1983). *Synergetics. An introduction. Nonequilibrium phase transitions and self-organization in physics, chemistry and biology.* Berlin: Springer.

Heiden, U. An der. (1992). Chaos in health and disease - phenomenology and theory. In W. Tschacher, G. Schiepek, & E. J. Brunner (Eds.), *Self-organization and clinical psychology. Empirical approaches to syngergetics in psychology* (p. 32-54). Berlin: Springer.

Horgan, J. (1995). From complexity to perplexity. *Scientific American*, *June*, 74-79.

Howell, D. (1997). *Statistical methods for psychology.* London: Duxbury.

Huff, D. (1986). *How to lie with statistics.* Harmondsworth: Penguin Books.

Kauffman, S. A. (1993). *The origins of order. Self-organization and selection in evolution.* Oxford: Oxford University Press.

Kelso, J. A. S. (1990). Phase transitions: foundations of behavior. In H. Haken & M. Stadler (Eds.), *Synergetics of cognition* (Vol. 45, p. 249-268). Berlin: Springer.

Kemeny, J. G., & Snell, J. L. (1976). *Finite marvov chains.* Berlin: Springer.

Kendall, S. M., & Ord, J. K. (1990). *Time series.* London: Edward Arnold.

Krauth, J., & Lienert, G. (1973). *Die Konfigurationsfrequenzanalyse (KFA) und ihre Anwendung in Psychologie und Medizin : ein multivariates nichtparametrisches Verfahren zur Aufdeckung von Typen und Syndromen.* Freiburg: K. Alber.

Krippendorff, K. (1986). *Information theory. Structural models for qualitative data.* London: Sage.

Lagasse, J., Courvoisier, M., & Richard, J.-P. (1977). *Logique séquentielle.* Paris: Dunod Universit.

Lemay, P. (1993). *The modelling of psychiatric patients' state evolution using two non-statistical methods* (Licence Research). University of Lausanne.

Lemay, P., Dauwalder, J.-P., Pomini, V., & Bersier, M. (1996). Quality of life - a dynamic perspective. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 276-295). Singapore: World Scientific.

Lemay, P., Vuadens, P., Dauwalder, J.-P., & Pomini, V. (1994). *Dynamique éco-systémique des comportements de tolérance* (Tech. Rep.). Université de Lausanne. Institut de psychologie.

Leussler, A., & Ham, P. van. (1979). Combinational systems. In R. Thomas (Ed.), *Kinetic logic. A boolean approach to the analysis of complex regulatory systems.* Berlin: Springer.

McCulloch, W. S., & Pitts, W. H. . (1974). A logical calculus of the ideas immanent in nervous activity. In W. Buckley (Ed.), *Modern systems research for the behavioral scientist* (p. xx). Chicago: Aldine Publishing Company.

Menard, S. (1991). *Longitudinal research.* London: Sage.

Morin, E. (1977). *La méthode. 1. La nature de la nature.* Paris: Seuil.

Nesselroade, J. R., & Ford, D. H. (1984). Methodological considerations in modeling living systems. In M. Ford & D. Ford (Eds.), *Humans as self-constructing living systems* (p. 47-79). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Ostrom, C. W. J. (1990). *Time series analysis. Regression techniques (2nd edition).* London: Sage.

Parry, A. (1996). Love is a strange attactor: therapy at the edge of chaos. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 235-255). Singapore: World Scientific.

Peitgen, H.-O., Jrgens, H., & Saupe, D. (1992). *Chaos and fractals. New frontier of science.* New York: Springer.

Pomini, V. (1997). *Auto-contrôle et réhabilitation.* Unpublished doctoral dissertation, University of Lausanne.

Pomini, V., Lemay, P., Dauwalder, J.-P., & Bersier, M. (1996). Dynamics of daily life. Interactions between individuals and their environment. *Swiss Journal of Psychology*, *55*(1), 49-60.

Rapp, P. E., Jiménez-Montano, M. A., Langs, R. J., Thomson, L., & Mees, A. I. (1991). Toward a quantitative characterization of patient-therapist communication. *Mathematical biosciences*, *105*, 207-227.

Schiepek, G., & Tschacher, W. (1992). Application of synergetics to clinical psychology. In W. Tschacher, G. Schiepek, & E. J. Brunner (Eds.), *Self-organization and clinical psychology. Empirical approaches to synergetics in psychology* (p. 3-31). Berlin: Springer.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Siegel, S., & Castellan Jr, N. J. (1988). *Nonparametric statistics for the behavioral sciences, 2nd edition.* London: McGraw-Hill.

Skinner, H. A. (1989). Butterfly wings flapping: do we need more 'chaos' in understanding addictions. *British Journal of Addiction*(84), 353-356.

Spiegel, M. R. (1980). *Probability and statistics.* New York: McGraw-Hill.

Thomas, R. (Ed.). (1979). *Kinetic logic. A boolean approach to the analysis of complex regulatory systems.* Berlin: Springer-Verlag.

Thomas, R., & D'Ary, R. (1990). *Biological feedback.* Boca Raton: CRC Press.

Tschacher, W., Scheier, C., & Grawe, K. (1998). Order and pattern formation in psychotherapy. *Nonlinear dynamics, psychology and life sciences*, *2*, 195-215.

Tufte, E. R. (1986). *The visual display of quantitative information.* Cheshire: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, Mass.: Addison-Wesley.

Wainer, H., & Thissen, D. (1993). Graphical data analysis. In G. Keron & C. Lewis (Eds.), *Data analysis in the behavioral sciences* (p. 391-457). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Waldrop, M. (1992). *Complexity: the emerging science at the edge of chaos.* New York: Simon and Schuster.

Ward, L. M. (1996). Chaos in psychophysics? Hypothesis testing and nonlinear forecasting approches. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 77-89). Singapore: World Scientific.

Wolfram, S. (1994). *Cellular automata and complexity.* New York: Addison-Wesley.

Zausner, T. (1996). The creative chaos: speculations on the connection between nonlinear dynamics and the creative process. In *Nonlinear dynamics in human behavior* (Vol. 5, p. 343-349). Singapore: World Scientific.