

Sequence analysis as a method for psychological research

JAN IVANOUW

Sequence analysis, based on the mathematical Markov model, is useful for analyzing interdependent data. Sequences can be recorded as event, state or timed event data, or by coding fixed time intervals. The probabilities of transitions in the data stream are used for modeling of the change processes. The first-order Markov model assumes that every event is influenced only by the immediately preceding event. The second-order model assumes influence from the two preceding events, etc. The models are characterized by their stationarity, the degree to which the same rules apply throughout the sequence. Statistical methods are introduced for testing of these characteristics as well as the homogeneity of sequences in a target sample. Lag sequence analysis is introduced as a method for simplifying analysis of higher order sequences. Advanced models, applications and software are introduced.

Correspondence concerning this article should be addressed to Jan Ivanouw, Nordborggade 6, 2100, Copenhagen Ø, Denmark, email: ivanouw@post8.tele.dk, tel: +45 35263225

This paper is an introduction to the method for data analysis called sequence analysis which is based on the mathematical Markov model. The sequences which can be described by this model, are called Markov processes. The method has wide applications and is for instance used in analysis of DNA-sequences in genome research. The present paper describes the method for use in the social sciences. For a comprehensive introduction consider Bakeman and Gottman (1977) or Gottman and Roy (1990). A more general review can be found in Abbott (1995).

Most statistical methods of data analysis in psychological research require that the observations are independent of each other. Often, though, this condition is not satisfied, and even in some instances the very object of research is the interdependence of events. If in these instances the researcher uses the traditional quantitative methods of analysis where all observations are aggregated (e.g. summed), faulty outcomes may result, especially in the sense of regularities being overlooked. In the world of sequence analysis, the traditional method is called 'marginal analysis' because it uses the total scores (marginals) in tables representing interactions between events.

As an illustration, let us think of an example from the realm of medicine where data usually are aggregated. Let us look at treatment errors, faults committed by doctors and other medical staff. In marginal analysis after defining what is meant by a treatment error, one would count the number of errors which have occurred

in specified time intervals and also record some of the circumstances cooccurring with each faulty treatment. Data typically would be tabulated using some of the circumstances as independent variables and fault counts as the dependent variable. Regression analysis may also be used in order to determine the relative importance of the independent variables for fault occurrences. Now, what is missing in this type of data analysis is the chain of events resulting in the errors.

As a simple, anecdotal example of the importance of event sequences in this area, consider the observations of Semmelweis in 1847 of puerperal fever (childbed fever) among women in the General Hospital in Vienna. The women who gave birth in one hospital clinic, which was also a teaching service for medical students, had a very high risk of puerperal fever, while the women in another clinic at the hospital, which also were used for instruction of midwives, had a low risk. Semmelweis noted that the physicians examining the women in the first clinic often had performed autopsies on dead bodies in the morning before attending the ward. He then proceeded to demand that the young physicians washed their hands after performing the autopsies and before examining the women. This helped, but the young doctors got angry with him. Missing the point, they resented the perceived insinuation of being unhygienic and told him that they certainly did wash their hands when rising in the morning. This happened, of course, before infectious microorganisms were discovered. Even without knowledge of the infection mechanism, though, the sequence of 'hand washing', 'autopsy', 'medical examination of woman' highlighted the possible cause of the medical error.

As an example of how sequence analysis might be useful for statistical analysis in psychology, we may think of transference interpretations in psychotherapy. Through the history of psychoanalysis the value of interpretation of negative and positive transference has been discussed. The orthodox view in psychoanalysis is that interpretation of transference is the central intervention of importance (Strachey, 1934). Research have, however, failed to confirm a link between the mere frequency of transference interpretations and positive outcome. There even seems to be an inverse relation between proportion of transference interpretations on the one hand and on the other patient-therapist alliance and therapeutic outcome (Bergin & Garfield, 1994). In another line of research, immediate effects of transference interpretations were studied. Luborsky et al. (1979), for instance, located the 250 words in patients' wordings before and after each transference interpretation made by the therapist and coded the patients responses on different variables. This method introduced sequence in the sense of targeting reactions before and after a criterion (transference interpretation). For data analysis, though, a traditional frequency analysis of response categories among a fixed amount of data (the 250 words before and after the interpretation) was performed. The results

of this study seemed to contradict the view that transference interpretations would lead to a decrease in patient transference. Speculatively, though, proper analysis of the interaction sequences before and after the transference interpretations may have added information to the study.

In a study which compared frequency analysis and analysis of response sequences McCullough (1991) studied brief dynamic therapies consisting of 27-53 weekly sessions. Therapist interventions were classified as one of the following: 'interpretations involving the patient and therapist' (= transference interpretation), 'interpretations involving the patient and a significant other', 'clarification', and 'other relationship interpretations'. Patient behavior was coded in one-minute periods as either affective or defensive. Outcome of therapy was measured as residual gain from start to termination of therapy. The average effect size was 0.95, which may be regarded as considerable. The data were first analyzed using traditional frequency analysis of occurrence per session of each specific patient and therapist category. Using this method of analysis, the result was that type of intervention did not correlate with outcome. Data were also defined in terms of the mean frequency of patient affective or defensive responding per minute in the three minutes after each therapist intervention. While the frequency of defensive responding totaled for the whole session did not correlate with outcome, McCullough et al. found that frequency of patient defensive responding in the three minutes following a therapist intervention in fact was significantly negatively correlated with improvement. They also found that the frequency of affective responding in the three-minute period following a therapist intervention was significantly positively correlated with improvement. When split up according to type of therapist intervention, they found, furthermore, that transference interpretation followed by patient affect had a significant positive correlation with outcome of therapy. But they also found that nontransference interpretations when followed by affect were positively correlated with outcome. Sequence analysis also showed that the interpretation interventions were almost twice as likely to provoke defensive responses as was clarification.

This study showed the higher power of taking sequences into consideration compared to using marginal frequency analysis. The authors did correlate type of sequence with an extraneous variable (outcome), but they did not analyze the characteristics of the sequences per se, as may be done using the methods to be introduced here. The study picked two-event sequences (therapist intervention followed by patient behavior), but did not address longer sequences. For instance, the issue of why and when the therapist offers different types of interpretation might be examined by looking at the interaction before the therapist intervention. In the study, it was not examined whether the same sequences were found

throughout the session, and it was not examined whether comparable sequences were found across patients.

Next we will look more closely at the statistical analysis of sequence. In the following treatment the mathematical demands on the reader will be kept to a minimum.

Markov chains

The Russian mathematician A. A. Markov (1856-1922) developed mathematical models for sequences with well defined dependencies. As mentioned, many statistical methods require independence between events (data), while in daily life everybody knows that events are linked to each other by numerous cause-effect links. Markov systematized and simplified these apparently very complex links. In his simplest model (the first-order model) every event is only influenced by the immediately preceding event. This means that at a given point in time the next event would be influenced by the present event, but it does not matter what determined the present event. In the more complicated second-order model, every event is influenced by the two preceding events only, and so on. These models have the property of a well defined memory mechanism. As such, they are well ordered alternatives both to the often unrealistic assumption of totally independent events on the one hand, and on the other hand a likewise unrealistic notion that everything depends on everything else and that statistical analysis of the interdependency among events therefore is so complicated that in practical terms it is not possible.

Let us as data for our treatment of sequence models begin with a simple Tolman-like learning experiment, a rat having to choose between two doors, Left and Right, for obtaining the food that might be behind one of the doors. Let us assume that the result of 21 continuous trials in such an experiment is L R R L R L R R R L R R R L R L R L R R R, while in another experiment the response sequence is L R L L R L R R R R R R L L L R R R R R¹. Both series consists of 7 choices (35%) of the left door, and 14 (65%) of the right door. Looking at the sequences it is easy to notice that in the first sequence L is always followed by R, which is not the case for the second response sequence. This difference is lost if one looks only at the frequency distributions.

Types of sequence data

The data in the learning experiment example are called *event data*, which are single codes in a sequence without overlap between codes. With another type, *state data*, the duration of observations is also recorded. Such data, thus, consist

of a stream of codes and the start and end time of each code. The beginning of a new code implies the end of the previous code. *Timed event data* are the most general data type. Codes are timed like state data, but the intervals for the codes may overlap. The first two data types may also be represented as timed event data. A fourth type of data, *interval data* are generated using a time interval, for instance 10 seconds. For each of the successive intervals one or more codes are assigned (for instance McCullough et al., 1991). This method is rather simple. On the other hand these data are often not very precise as changes during an interval are not represented in the data. Because of the simplicity, this data type has been used extensively, but with the recent development of better equipment more sophisticated data registration will often be possible.

In some instances repeating codes are allowed. If an event is defined as a code with a duration lasting until the shift into another code, naturally codes cannot repeat. In interval data, the same code may be present in successive intervals. As an example, in an investigation of the behavior of couples during a 15-minutes videotaped discussion about an identified marital conflict (Gottman, Markman & Notarius, 1977), both speaker content, speaker affect and listener's nonverbal behavior was coded. A new behavioral unit was defined as starting when either speaker affect, content or listener affect changed. This means that even if the speaker talked about the same content with the same expressed affect, a new behavioral unit was coded if the listener changed affect, resulting in repeating codes for the speakers behavior. Slightly different statistical methods must be used for sequences with and without the possibility of repeating codes, although this will not be commented further in this treatment (a treatment of the problem of repeating codes may be found in Bakeman and Quera, 1995a). The different data types described are manifested as different ways of entering codes. On the other hand, the basic data process is common for all data types. The data stream is divided into distinct data points. For event data, one code is assigned to each data point. For state data, a small basic time unit may be defined as the basic data point. An event code will be considered as a repeating code at each of the data points in the time span entered for this code. For timed event data, each of the data points may contain several codes.

The transition matrix

For making sequences amenable to analysis, one can imagine a window moving along the sequence allowing the spectator to see consecutive pairs of codes. Using as an example the first of the learning sequences mentioned above, L R R L R L R R R L R R R L R L R L R R R, first one would see 'LR' in the window. Next, after moving the window one step, one would see 'RR', then 'RL' and so on.

Notice that the coding pairs are overlapping, so the last code from the previous pair becomes the first in the next pair. All the pairs are tabulated in table 1.

Table 1. *Transition frequencies of the first learning sequence example*

frequency	After		Total (rate)
	L	R	
Before			
L	0	7	7 (0.35)
R	6	7	13 (0.65)
Sum	6	14	20 (1.00)

Because each code is counted twice, except the edge codes, the overall table total is one less (20) than the sum of the individual codes (21). Repeating codes are allowed, and in fact there are seven instances of RR. Calculating row percentages one gets a *transition probability table* (table 2).

Table 2. *Transition probabilities of first learning sequence*

frequency	After		Total
	L	R	
Before			
L	0.00	1.00	1.00
R	0.46	0.54	1.00

Generalizing from this table, it is possible to predict what will happen after a given code. In this instance, always when L occurs one will expect R to follow. When R occurs one would expect the following code to be L with 0.46 probability and to be R with 0.54 probability. The second learning sequence example (L R L L R L R R R R R L L L R R R R R) with the same distribution of single codes gives the transition probability table 3.

Table 3. *Transition probabilities of the second learning sequence example*

frequency	After		Sum
	L	R	
Before			
L	0.43 (3)	0.57 (4)	1.00 (7)
R	0.23 (3)	0.77 (10)	1.00 (13)

Note transition frequencies are in parenthesis.

In this table, one will note an approximately equal probability of L being followed by L and R, while the code following an R has a .77 probability of being another R and only .23 probability of being L. The identical high total number of Rs (65%) in the two sequences seems, though, to have different causes. In the first, an L

leads inevitably to an R. In the second, there is not such a tendency. There is an equal probability of getting the two codes after an L. In the second instance, it is the R that mainly produces another R. We have, thus, two different processes leading to the same total (marginal) result. Only by using sequence analysis it is possible to demonstrate the different processes.

The transition table resembles the familiar crosstable, but there is an important difference. In the transition table the data points are not independent, as it is the relations between a given (before) and target (next) code which is tabulated².

Analyzing the transition matrix

In transition tables, as in the above examples, it is important to look for regularities, like the possibility of predicting the next code from a preceding code. Testing for regularities in the table, one applies a statistical test of the null hypothesis that there are no such regularities, which is equivalent to testing for independence in the table. Several statistical tests are possible (Gottman & Roy, 1990). The familiar chi square-test can be used, but there are some advantages of using another type of chi square-test, based on likelihood ratio, which is often called the G^2 -test (Agresti, 1996). If the analysis rejects the null hypothesis, it means that there is in fact some regularity in the transition table and hence in the sequence. Then, one may look for the specific transitions that have significant higher transition probability than expected from the null hypothesis. This may be done by calculating the residuals for each of the cells in the table. These show how much the individual cells in the table deviate from the independence assumption. Several types of residuals can be calculated (Bakeman & Gottman, 1997), but the general idea is to calculate a type of standard scores, z-scores. Individual cells with z-scores numerically greater than 1.96 are regarded as statistically significant at the 5% level, and accordingly indicate cells with transitions of special interest.

In the first of the learning examples above, the likelihood-ratio chi-Square is 6.5 and the approximative p-value is 0.01. This shows that the table is not random, but that there is some regularity in the table. Another way to state this, is to say that the statistical test compares the zero-order (independence) and the first-order Markov model of dependence between a code and the next code. In other words, at least a first-order Markov model is appropriate to describe the sequence. The z-score residuals can be seen from the table 4. (In a four cell table all residuals will be numerically equal and as such just indicate the same as the chi-square test).

Table 4. *Residual z-scores from first learning sequence*

z-score Before	After	
	L	R
L	-2.15	2.15
R	2.15	-2.15

In the second of the learning examples above, the likelihood-ratio chi-Square is 0.8 and the approximative p-value is 0.36. This shows that, although there seems to be a regularity in the table, it is not possible to reject the null hypothesis that the observed regularity is just a random variation. In other words, we can not reject the zero-order Markov model. Accordingly, as can be seen in table 5, no z-score residuals are numerically greater than 1.96.

If there had been more than two possible codes for each event, the table would have had more rows and columns, and even if the overall chi-square test was insignificant (no regularity in the table), some cells *might* contain z-score residuals greater than 1.96. These should not, however, be trusted as they might just be type I errors (see below).

Table 5. *Residual z-scores from second learning sequence*

z-score Before	After	
	L	R
L	0.92	-0.92
R	-0.92	0.92

Characteristics of a Markov sequence: Stationarity

When using a transition probability table for uncovering the structure of a sequence, you assume that the same transition probabilities apply to the whole sequence, which means that the same rules govern throughout the process. If this is true, the sequence is said to be *stationary*.

To test for stationarity the sequence in question is divided into subsequences, often two are used. You can use the log likelihood chi-square G^2 -test to see if the data divided into the subsegments are similar to the data without such division. This is the null hypothesis. A significant G^2 -test would point to nonstationarity. The division into segments could be random, and for testing stationarity it may be appropriate to test several different random divisions.

The sequence could also be divided according to a hypothesis about the existence of different subphases. For instance, Gottman et al. (1977) divided

the analysis of the earlier mentioned conflict discussion into three subsegments (presenting the problem, argue about solutions, and reaching compromise). In the Gottman example, the whole sequence of the conflict discussion was non-stationary, but the sequence in each of the subphases had stationarity. This means that the researchers had identified different characteristic interaction sequence codes which were used consistently throughout each subsegment.

One interesting use of sequence analysis is to compare interaction over time, for instance before and after a psychological intervention. The goal in this application would be to test for difference of sequences, non-stationarity, across the intervention with the hypothesis of finding stationarity in the sequences before and after, respectively.

The psychological concept of development implies that 'the rules change' over time. Berchtold and Sackett (2002) worked with young monkeys. When analyzed without consideration of development the behavior sequences of the monkeys appeared to be non-stationary, but when subdivided according to development stages, the behavior sequences were stationary. In order to be able to describe the rules governing the change in transition probabilities from one phase to the next, the researchers generated what they called the Double-Chain Markov Model (DCMM). The point is to develop rules that can predict when each of the primary transition matrices (rules) apply, and how one set of rules replaces another. For this purpose they developed a probability transition matrix that describes the transitions from one to another of the primary transition matrices.

Characteristics of a Markov sequence: Homogeneity

Often, sequence analyses have been performed with a single case (person, monkey, couple) which, if observed for a sufficiently long duration, will provide enough data to perform some statistical analyses of the sequence. Single case analysis have often been used as an exploratory method for charting new research areas and developing hypotheses. However, the aim of research typically is to find regularities that hold for all individuals of the goal population (or pertaining to subgroups of the population). In order to do this, it is necessary to compare sequences across cases to determine if the sample is homogenous in regard to the sequences of interest. If this is the case, the data can be aggregated for the whole sample so the greater amount of data will provide more power to analyses and the testing of hypotheses.

For analysis of homogeneity, the log likelihood chi-square G^2 - test can again be used. A significant result for the whole sample would mean that there is some degree of inhomogeneity. If the sample is not sufficiently homogeneous, it may be a research goal to establish if there exist subgroups in the sample, each of which

is homogeneous. The sample may be divided according to different hypotheses about subgroups and influencing variables. These divisions can again be tested with the G^2 -test, and the results of different ways to divide the sample can be compared using statistical criteria for model selection (see for instance Agresti, 2002, p. 216ff or Harrell, 2001, p. 202ff).

As an example, Krokoff et al. (1988; Gottman and Roy, 1990) found that the interaction sequences in the conflict discussions of 120 couples showed inhomogeneity for the group as a whole. They then analyzed the sample subdivided as to state of satisfaction with the marriage (happy versus unhappy), social class (white-collar versus blue-collar) and conflict tactics (conflict engagers versus conflict avoiders). Analyzed within the subgroups resulting from combining these variables, the sample showed markedly less inhomogeneity.

Characteristics of a Markov sequence: Order

In the examples used up till now it was hypothesized that the rules of the sequence, if any at all, pertained to the connection between an event and the next event, that is the first-order Markov model. There are, though, situations in which more than one event might influence what comes next. An example would be patient behavior after a transference interpretation. If the patient is negative after the therapist's intervention, it might be the therapist intervention that influenced the patient to be negative. But what if the patient started out as being negative, and the therapist in response to this negativity intervened with the transference interpretation, and in case the interpretation did not influence the patient, he or she just continued to be negative? In this example, the patient's negativity is best understood as a continuance of his or her own behavior before the therapist intervention, which means that it has to be seen in relation to a state lying two steps back.

In sequence analysis, the concept 'lag' describes how many events lie between two events in question. Lag 0 is the starting event. Lag 1 means the relation between two adjacent events. In the above example, analyzing the patient's behavior in relation to his own behavior before the therapist intervention would be to look at the sequence at lag 2. A sequence where there is dependency between events at lag 1, but not at lag 2, is a sequence of the first order. A second-order Markov model implies that events at lag 2 have to be included in the analysis to understand the interdependencies in the sequence. In the therapy example, by looking at lag 2 data one might find patient negativity followed by therapist interpretation followed by continuing patient negativity, and thus identify a sequence where therapist intervention did not change patient affect.

In the metaphor of the moving window, for a second-order sequence the window is broader so you can see three codes simultaneously. In the first learning example, the sequences of three codes are (with counts in parenthesis) LLL (0), LLR (0), LRL (3), LRR (4), RLL (0), RLR (6), RRL (3) and RRR (3). These data can be tabulated in a set of tables. When there are two possible codes for each event, you will need two tables, only (table 6).

Table 6. *Three-event transition probabilities for first learning sequence*

First		3. code		Total
L	2. code	L	R	
	L	0.00 (0)	0.00 (0)	0.00 (0)
	R	0.43 (3)	0.57 (4)	1.00 (7)
	Total	0.43 (3)	0.57 (4)	1.00 (7)
R	2. code	L	R	
	L	0.00 (0)	1.00 (6)	1.00 (6)
	R	0.50 (3)	0.50 (3)	0.00 (6)
	Total	0.25 (3)	0.75 (9)	1.00 (12)

Note transition frequencies are in parenthesis.

The table can be read as follows: If the first code was L and the second was R, then the probability of the third being L is 0.43, while the probability of the third being R is 0.57. It is possible to test the implied second-order model against the first-order, meaning if it really is necessary to include events at lag 2. The likelihood ratio chi-square G^2 - test again could be used to compare the two models, testing the second-order against the first-order model. If the test is significant, at least the second-order model is necessary to describe the sequence data - in fact, higher order models may be considered. If the test is not significant, the first-order is sufficient (given the fact that the test of the first-order model against the zero-order model was found to be significant).

It is readily apparent that if there are more than two possible codes for each event, and if we want to check the influence from several more lags, say also lag 4 and 5, the transition tables will have many cells, and it would require a huge amount of data to get reliable estimates of probabilities and to have enough power for the statistical testing.

Let us say that, when analyzing a possible second-order model, we just ignore the second (intermediate) code. Then, we can simplify the table using the column totals for the third code in table 6 and we will get the probability transition table 7.

Table 7. Lag 2 transition probabilities for the first learning sequence

1.code	3. code		Total
	L	R	
L	0.43 (3)	0.57 (4)	1.00 (7)
R	0.25 (3)	0.75 (9)	1.00 (12)

Note transition frequencies are in parenthesis.

In this collapsed table of lag 2 transitions, the likelihood-ratio chi-Square is 0.6 and the approximative p-value is 0.42, which means that we cannot reject the null hypothesis. This shows that although for this sequence we found a regularity of first order, meaning a link from a code to the following code, there is no dependence at lag 2 if we ignore the intervening code. In other words, the appropriate Markov model seems to be of the first order. This principle is used in lag sequential analysis.

Lag sequential analysis

One way of circumventing the problem of very complex sets of transition probability tables for higher-order Markov chains is to use lag sequential analysis. In this method, in principle, one transition table is produced at each lag. A specific code is selected as the criterion (at lag 0), and transition probabilities of all the other codes are calculated with respect to the criterion code as a function of lag from the criterion. That means a table linking the selected code at lag 0 and all the codes in question at lag 1, another table linking the code at lag 0 and all other codes at lag 2, a third table linking the code at lag 0 and the possible codes at lag 3, and so on. For practical reasons, such tables are usually combined into a single table (an example of which is table 8).

In general, the important sequences in a lag sequence table are identified through three steps. First, say we are looking into the possible codes A, B and C. The code with the highest lag 1 conditional probability from the criterion behavior A may be the code C. At lag 2 the code with the highest lag 2 probability from the criterion A may be the code B. A probable sequence A - C - B is, thus, identified. The second step in identifying a sequence is to note that this would be a likely sequence only if the lag 1 transitional probability of behavior C being followed by behavior B also showed a peak probability. Third, the z-scores for testing the difference between the actual transition probabilities and those expected under the null hypothesis have to be greater than 1.96. Otherwise, the transition probabilities can not be considered as different from the base rate, and in that case we have not confirmed the sequence A - C - B.

Table 8. Lag sequence analysis of conflict discussion data from satisfied and unsatisfied married couples

HPFo		1	2	3	4	5	6
satisfied couples	WPFo	0.24	0.13	0.18	0.16	0.17	0.18
	WAGo	0.30*	0.10	0.19*	0.10	0.15*	0.11
	HPFo	0.00	0.38*	0.14	0.26*	0.17	0.22
	HAGo	0.02	0.11	0.07	0.10	0.08	0.10
	z-score ^a	9.77	10.3	5.11	3.57	2.38	-
non-satisfied couples	WPFo	0.23*	0.11	0.17*	0.14	0.16*	0.12
	WAGo	0.16	0.00	0.08	0.10	0.06	0.10
	HPFo	0.00	0.33*	0.15	0.25*	0.17	0.20*
	HAGo	0.01	0.10	0.05	0.10	0.06	0.10
	z-score ^a	3.90	9.11	2.50	5.02	1.86	2.10

Note From "The topography of marital conflict: A sequential analysis of verbal and non-verbal behavior," by J. Gottman, H. Markman and C. Notarius, 1977, *Journal of Marriage and the Family*, 39, p. 471. Copyright 1977 by National Conference on Family Relations. [Reprinted with permission]. The asterisks (*) show transition probabilities significantly different from the null hypothesis.

^a The z-score is for the same column transition probability marked with an asterisk (*).

Let us again look at the study of conflict discussion in marital couples (Gottman, Markman & Notarius, 1977). The interactions were coded for content. Among the codes were 'statements about a problem including expressing feeling about the problem' (PF) and 'expressing agreement with the partner' (AG). This could be done by the husband (H) and by the wife (W). Besides, the affective tone was recorded as negative (-), neutral (o) or positive (+). The sample was divided into a group of couples satisfied with their marriage and a group of couples not very satisfied. Some data from the study are presented in table 8.

The transitional probabilities are computed, first, for transition from the given code to the following, target code (lag 1). The lag 0 code in question may be HPFo (husband talks in a neutral way to his wife about a problem). When looking only at codes with neutral effect, the lag 1 codes may be WPFo, WAGo, HPFo or HAGo. One of the cells, for instance, contains the transition HPFo-WAGo (husband neutrally stating problem followed by wife neutral agreement). At the next lag, the transition probability from the given code to the next following code, ignoring the intervening code, is computed (lag 2). One of the cells contains the sequence HPFo-X-WAGo, where X can be any of the possible codes, and X is not tabulated. For lag 3, the transitional probability from the given to the third code, ignoring the two intervening codes, is computed. One of the cells contains HPFo-X-X-WAGo, etc. Thus, when looking at one row in the table we can examine the

transitional probabilities for one particular given code into one particular target code (for instance HPFo to WAGo) at different lags³. We are spared a lot of detail because we do not differentiate between the possible codes which may take the place of X

In the example, the researchers were interested in what happens after one of the spouses stated a problem in a neutral way (HPFo or WPFo) and excluded interaction with negative or positive affect. As described, they tabulated the responses following the problem statement as either problem statement/feeling or agreement (here is shown only the table for interactions starting with the husband's problem statement/feeling). As mentioned, it was possible for codes to repeat (say a HPFo may be followed by another HPFo code).

Looking at the satisfied couples in table 8, the code with the highest lag 1 conditional probability from the criterion behavior HPFo is the code WAGo (.30). The z-score for this transition is 9.77, way over the 1.96-criterion. Next, the code with the highest lag 2 conditions probability from the criterion is the code HPFo (.38) with the z-score 10.31. Next again, the (lag 3) code following the starting criterion HPFo is WAGo (.19) with the z-score 5.11. At lag 4, the highest transition probability is HPFo (.26) with the z-score 3.57, and at lag 5, the highest transition probability is WAGo (.15) with the z-score 2.38. At lag 6, the highest transition probability is HPFo (.22), but the z-score is lower than 1.96 (not entered in the table). We have, thus, found a probable sequence HPFo - WAGo - HPFo - WAGo - HPFo - WAGo. This is the type of interaction couples therapists often call 'validation': when one partner describes a problem, the other listens and confirms what has been said. On the contrary, for the not satisfied couples one can in the table identify another possible sequence: HPFo - WPFo - HPFo - WPFo - HPFo - WPFo - HPFo, which may be called 'crosscomplaining'. Each partner in turn responds to the partner's problem statements with his or her own problem statement.

The second step of the lag sequential analysis, controlling for the transition probabilities for the codes in the possible sequence in turn starting with the codes following the starting criterion, was reported in the Gottman et al. paper for the code WPFo (not reproduced here) which showed the expected dependencies among the non-satisfied couples. For a test of the sequence for the satisfied couples where the code following HPFo was WAGo, WAGo should be treated as the given, lag 0, code, with HPFo as the next code. This analysis was not shown in the Gottman et al. paper.

Because lag sequential analysis is based on selection of special data transitions of interest, the method is vulnerable to type I error. This is a problem that may occur when many statistical tests are performed. The resulting p-value represents the probability of committing a type I error⁴ when rejecting the zero hypothesis. Testing for significance 100 times using a 5% criterion would, thus, yield a

mean of 5 erroneously significant results. When examining significant transition probabilities in the cells of an analysis with many tables, the probability of type I error rises significantly. One way of containing this problem is to start with an omnibus test of the complete set of data tables, and only if this test shows that statistical significant regularities may be found in the tables, post hoc analyses, as those demonstrated above, are suitable. Several possibilities for such testing is described by Bakeman & Gottman (1997).

Another technique for analyzing higher-order Markov chains is *log linear analysis*. This is a statistical method often used for multi-dimensional tables. One asset of this method is that it can model data without the simplification inherent in the lag sequential method, where intervening codes are ignored. Everitt (1996) gives an introduction to log linear analysis, while a general treatment of can be found in Agresti (1996). Gottman and Roy (1990) demonstrate how the method can be used for sequential analysis.

Advanced models

As statistical techniques develop, more advanced models will be available. Allcroft, et al. (2004) describe different statistical models with different amounts of memory. They mention the hidden Markov model (HMM) as a model for the hypothesis that a series of unobserved underlying states follow a Markov chain, and that the observed behaviors are independent of each other but dependent on the underlying states. Another model, the latent Gaussian model, assumes that events occur when they exceed threshold of an underlying continuous variable, thereby incorporating some 'short-term memory'. A third model, the semi-Markov model, uses the duration of time spent in the previous state as a co-influence on the next state. This model can, thus, be said to incorporate 'longer-term memory'.

It has to be remembered, though, that in research it is an asset to be able to identify the simplest model that explain the relevant data. If, for instance, a non-sequential (marginal) model fits the data, it is preferable to a sequence model, and among these models the one with the least parameters, the model of lowest order, is preferable. Also, more advanced models are only relevant in so far as they enhance the understanding of the processes in question.

Applications

Throughout this introduction, examples have been drawn from medical treatment procedures, learning processes, and interaction in therapy and among married couples. Benjamin (1979) introduced sequence analysis as a method using her

SASB interaction coding scheme on therapist-patient interaction. Gottman and Roy (1990) present many examples mostly concerning communication. A few other examples of research using sequence analysis follows.

Allcroft, et al. (2004) used different statistical models for describing animal feeding behavior. Social behavior among animals was examined by Jensen and Wood-Gush (1984). Arousal and the behavior states of premature infants were examined by Giganti, et al. (2002), and the interaction of adults and infants was modeled by Fogel (1982). Hu & Shapley (2003) used the Markov chain as a model for authority relations. Jeong (2006) examined conversational language in discussions in on-line group debates. The communication between doctors and patients has been examined in several studies (for instance Eide, et al., 2004; Kim, et al., 2005; Salmon, et al., 2006). In the clinical realm, expressed emotion and relationships in disturbed families was examined by Cook, et al. (1989). Bus and Kruijenga (1989) studied the problem-solving behavior among experts while diagnosing reading problems

Software

The analysis described in this paper is greatly helped using PC-programs. General purpose statistical programs, like SPSS, SAS or S-PLUS, may be used. But the conversion of raw data into data sequences amenable for sequence analysis can be tedious. This process is greatly helped by use of a dedicated sequence analysis program like GSEQ (Bakeman & Quera, 1995b). This program is cheap and rather simple in use. It is useful for the first steps in sequence data analyses, and for further analyses it can export data for use with general statistical programs, like SPSS. Another PC-program, MARCH, is more complex and includes analysis of different types of Markovian models, like the hidden Markov model and the double chain Markov model (Berchtold, 2001).

REFERENCES

- Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, 21, 93-113.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.
- Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.
- Allcroft, D.J., Tolkamp, B.J., Glasbey, C.A. & Kyriazakis, I. (2004). The importance of 'memory' in statistical models for animal feeding behaviour. *Behavioural Processes*, 67, 99- 109.
- Bakeman, R. & Gottman, J.M. (1997). *Observing Interaction. An introduction to Sequential analysis*. Cambridge: Cambridge University Press.
- Bakeman, R. & Quera, V. (1995a). Log-linear approaches to lag-sequential analysis when consecutive codes may and cannot repeat. *Psychological Bulletin*, 118, 272-284.
- Bakeman, R. & Quera, V. (1995b). *Analyzing interaction. Sequential analysis with SDIS and GSEQ*. Cambridge: Cambridge University Press.

- Benjamin, L.S. (1979). Use of Structural Analysis of Social Behavior (SASB) and Markov chains to study dyadic interactions. *Journal of Abnormal Psychology, 88*(3), 303-319.
- Berchtold A. (2001). Markov chains computation for homogeneous and non-homogeneous data: MARCH 1.1 user's guide. *Journal of Statistical Software, 6*, 1-82.
- Berchtold, A & Sackett, G. (2002). Markovian models for the developmental study of social behavior. *American Journal of Primatology, 58*, 149-167.
- Bergin, A.E. & Garfield, S.L. (1994). *Handbook of psychotherapy and behavior change*. New York: Wiley.
- Bus, A. G. & Kruizenga, T. H. (1989). Diagnostic problem-solving behavior of expert practitioners in the field of learning disabilities. *Journal of School Psychology, 27*(3), 277-287.
- Cook, W. L., Strachan, A. M., Goldstein, M. J. & Miklowitz, D. J. (1989). Expressed emotion and reciprocal affective relationships in families of disturbed adolescents. *Family Process, 28*(3), 337-348.
- Eide, H., Quera, V., Graugaard, P. & Finset, A. (2004). Physician-patient dialogue surrounding patients' expression of concern: applying sequence analysis to RIAS. *Social Science and Medicine, 59*(1), 145-155.
- Everitt. B. S. (1996) *Making sense of statistics in psychology*. Oxford: Oxford University Press.
- Fogel, A. (1982). Early adult-infant interaction: Expectable sequences of behavior. *Journal of Pediatric Psychology, 7*(1), 1-22.
- Giganti, F., Hayes, M. J., Akilesh, M. R. & Salzarulo, P. (2002). Yawning and behavioral states in premature infants. *Developmental Psychobiology, 41*(3), 289-296.
- Gottman, J., Markman, H. & Notarius, C. (1977). The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior. *Journal of Marriage and the Family, 39*(3), 461-477.
- Gottman, J.M. & Roy, A.K. (1990). *Sequential Analysis*. Cambridge: Cambridge University press.
- Strachey, J.(1934) The nature of therapeutic action in psychoanalysis, *International Journal of Psychoanalysis, 15*, 127-159.
- Harrell, F. (2001). *Regression Modeling Strategies*. New York: Springer.
- Hu, X & Shapley, L.S. (2003). On authority distributions in organizations: equilibrium. *Games and Economic Behavior, 45*, 132-152.
- Jensen, P. & Wood-Gush, D. G. (1984). Social interactions in a group of free-ranging sows. *Applied Animal Behaviour Science, 12*(4), 327-337.
- Jeong, A. C. (2006). The effects of conversational language on group interaction and group performance in computer-supported collaborative argumentation. *Instructional Science, 34*(5), 367-397.
- Kim, Y.M., Kols, A., Prammawat, S. & Rinehart, W. (2005). Sequence Analysis: Responsiveness of Doctors to Patient Cues during Family Planning Consultations in Mexico. *Patient Education and Counseling, 58*(1), 114-117.
- Krokoff, L. J., Gottman, J. M. & Roy, A. K. (1988). Blue collar and white collar marital interaction and communication orientation. *Journal of Social and Personal Relationships, 5*, 201-221.
- Salmon, P., Humphris, G. M., Ring, A., Davies, J. C. & Dowrick, C. F. (2006). Why Do Primary Care Physicians Propose Medical Care to Patients With Medically Unexplained Symptoms? A New Method of Sequence Analysis to Test Theories of Patient Pressure. *Psychosomatic Medicine, 68*(4), 570-577.

NOTES

- 1 For reasons of presentation these examples are very short. In practical work you will need longer sequences
- 2 One anonymous reviewer commented that this dependency only is a problem when using exact statistical tests. When using asymptotic statistical tests, transition tables may be analyzed like other frequency tables. However, this requires a greater amount of data than is present in the actual example
- 3 Since these transition probabilities are 'taken out' of the not included complete tables of transition probabilities, the entries in the lag sequence table does not sum to 100%
- 4 While a type I error is the acceptance of a hypothesis that is false, a type II error is the rejection of a hypothesis that is true.