DOCUMENT RESUME

ED 053 813 PS 004 918

AUTHOR	Freedle, Roy; Lewis, Michael
TITLE	Application of Markov Processes to the Concept of State.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, Md.
REPORT NO	ETS-RB-71-34
PUB DATE	71
NOTE	15p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Environmental Research, Individual Differences, *Infant Behavior, *Parent Child Relationship, *Research Methodology, Tables (Data), *Verbal Communication
IDENTIFIERS	Markov Process

ABSTRACT

The purpose of this paper is to outline some application of the Markov Process to the study of state and state changes. The essence of this mathematical concept consists of the analysis of sequences of infant responses in interaction with its environment. Categories can be defined which reflect the joint occurrence of an infant's behavior (or condition) along with some associative event(s) in the infant's immediate environment. Each of these categories of infant-environment interaction can be used as a definition of state for the purposes of studying the sequential unfolding among categories. An example utilizing child vocalization data collected by Lewis is given. When applied to mother-infant interaction, a particular mother-infant pair may yield data which give a poor fit in terms of matching statistics with the Markov model. Therefore, three alternative procedures are suggested. (Author/WY)

Application of Markov Processes to the Concept of State[1]

Roy Freedle and Michael Lewis

Educational Testing Service
Princeton, New Jersey

The purpose of this paper is to outline some applications of a mathematical concept known as a Markov Process to the study of state and state changes. The essense of this mathematical concept consists of the analysis of sequences of infant responses in interaction with its environment. Categories can be defined which reflect the joint occurrence of an infant's behavior (or condition) along with some associative event(s) in the infant's immediate environment. Each of these categories of infant-environment interaction can be used as a definition of <u>state</u> for the purposes of studying the sequential unfolding among these categories.

The vocalization data collected by Lewis (this volume) will be used to provide an example. The vocalization data can be categorized into six states: neither mother nor infant vocalize; infant vocalizes alone; mother vocalizes alone to infant; mother vocalizes alone to some other person; mother and infant both vocalize; and mother vocalizes to another person and the infant vocalizes. We shall designate these six categories by the symbols 0, 1, 2i, 2, 3i and 3, respectively. Notice that the six categories (or states) clearly reflect the possible interactions that can occur between the infant and its immediate environment as far as vocalization behavior of both is concerned.[2] Furthermore, since the data were collected every 10 sec. for a total of 720 successive 10 sec. periods, exactly one of these six states can be said to have occurred for each interval.

A Markov analysis of these sequences of vocalization states will now be illustrated. The key concept that is needed is that if one knows which of the six states has occurred on trial n-1 that this is all one needs to know in order to adequately predict (using conditional probabilities) what state will occur in the next 10 sec. interval (this next interval can be designated as trial n). To be more specific, this Markovian assumption asserts that it would be irrelevant to the adequacy of our predictions if we also knew the state of the vocalization system that had occurred on any of the previous trials such as trial n-2, trial n-3, and so on.[3] All one needs for purposes of making adequate predictions about the sequences of states is a knowledge of what state occurred on the immediately preceding trial, trial n-1. When a sequence of data satisfy this assumption it is said to have a Markovian property.[4]

In order to test the adequacy of this Markovian assumption for the vocalization behavior and to allow for individual differences to emerge across each mother-infant pair, it is necessary to estimate the transitional probabilities (the conditional probabilities) using the raw data. A separate transition table of probabilities will be found for each mother-infant pair in the example.

It is easy to show how the transitional probabilities can be estimated. Consider the following succession of states obtained from coding the successive 10 sec. periods for a particular mother-infant pair: 3, 0, 1, 3, 1, 2, ... Set up a matrix with six rows and six columns labeled 0, 1, 2i, 2, 3i and 3, reading from the top down for the rows and similarly labeled reading from left to right across the columns. Using the above sequence, notice that the first pair of states is 3,0. Enter a tally in the rows labeled '3' and the column marked '0'. The second and third states form the next pair of states which is 0,1. Enter another tally in the row labeled '0' and the column labeled '1'. The next pair of states is 1,3; so enter a tally in row '1' and column '3' and so on until all successive pairs of states have been tallied. When this is done, sum up the tallies for each row and divide the frequency in each row cell by the sum for that row. These proportions that result in each row are then used as the estimates for the transitional probabilities of the transition matrix.

For the data under consideration here, there were 719 tallied entries for each mother-infant studied. The result of this tallying process for one such mother-infant pair is given in Table 1.
In order to test the adequacy of the Markovian assumption made above, a simulation of the successive vocalization states was carried out using the transition probabilities. This simulation technique relies solely upon the use of a random number table in conjunction with the transition matrix values. In order to initiate this simulation process one needs to select one of the six states as the starting state. This can be done in the following way. Since the frequency with which tallies occurred in each of the six rows have already been found, one sums all of the six frequencies and divides each row sum by this overall frequency (for the present data this overall frequency would be 719). This generates an <u>initial probability vector</u>: it tells one the likelihood that the mother-infant vocalization interaction would be in any one of the six states if any one of the 719 time intervals was picked at random. The

simulation process begins by selecting any one of these six states. In other words the only use of the initial probability vector is to initiate the simulation process. Once it is started one uses

Insert Table 1 about here

only the transitional probability matrix and the random number table to generate all further sequences of states. Once this simulation has been completed for each mother-infant pair studied, the assertion is made that any statistical manipulation of the real data that can be made can also be made with the simulated data; and furthermore, the two calculations should be virtually identical (i.e., statistically they should be indistinguishable from each other). In other words, the adequacy of the Markovian model to capture the essential characteristics of the data is assessed primarily in terms of how well it matches the same set of statistics that are calculated using just the raw data.[5] For example, one statistic of interest in the raw data is the number of successive 10 sec. periods that the infant alone vocalizes (category '1'), another interesting statistic is the number of successive 10 sec. periods that the mother alone vocalizes (category '2'), and our final example of an interesting statistic is the number of successive periods that both mother and infant vocalize in the same time period (category '3').

For our first mother-infant pair studied these observed calculations for these three statistics based on the raw data are presented in Table 2; a similar calculation was performed using the simulated data based on the application of the Markov model. The results of the simulation for each of these three statistics are also given in Table 2. One sees that the Markov model appears to capture much of the detailed characteristics of the data for this particular mother-infant pair. The infant in this case was female; this example was chosen because both mother and infant were the most vocal of all mother-infant pairs studied.

Insert Table 2 about here

Another subject (male infant) was selected for a similar type of analysis because it was very low in overall frequency with which the infant vocalized--the mother in this case was also very low in vocalization behavior.

The transitional probability matrix was analyzed in the same way as the previous mother-infant pair. This matrix is given in Table 3. The adequacy of the Markovian assumption to capture the essential details of the succession of vocalization states was again assessed by determining the number of successive 10 sec. periods of states '1', '2', and '3' just as in the previous example. The observed and predicted (simulated) frequencies are presented in Table 4. Again observe a close match in frequencies by use of the Markov model simulation.

Insert Tables 3 and 4 about here

<u>Use of Markov Model</u>

There are several ways such a modeling can be of value in the study of interaction. Observation of the transitional probability matrix (for each mother-infant pair) reveals the ways in which a current state of the mother-infant system influences the conditional probability of the next state (see Tables 1 and 3). By examining the magnitude of the diagonal probabilities, one can get an immediate estimate of the degree to which the infant will persist in state '1', one can get an estimate of the degree to which the mother and infant together will persist in state '3', and so on.

For the female infant the largest conditional probability in each row is generally along the main diagonal of the matrix (see Table 1). This indicates that state tends to persist over time--that is, this subject will tend to have many long runs of the same state. The male infant's data indicate, however, that most states usually revert back to the '0' category. This reflects the differences between these two infants in the amount of vocalization both they and their mothers exhibit.

Consider other individual differences. Let us take each row of the transition matrix which applies to the female infant and compare it with the entries for the same row of the matrix for the male infant. For the female infant we see that if the previous state of the vocalization system was '0' the state of the system in the next time interval will again be state '0' (probability .42). The next most frequent state following a '0' is state '2' which occurs with probability .22. For the male infant there is again a large probability that if the vocalization system was in state '0' in the previous time interval it will persist in this state for the next time interval (probability .77). The next most frequent state is state '2'; however, this occurs with a very low probability of .08. Hence for this first row both the male and female transition matrices are similar in the sense that the most probable event following '0' is the persistence of the same state '0' while the second most probable event following '0' is state '2'. The male and female transition matrices differ, though, in the magnitude of these entries.

For the female infant the most likely state following state '1' is a persistence of the same state (probability .46); the second most likely event following state '1' is a tie between moving to state '0' and moving to state '3'. The male infant shows a different pattern. The most likely event following state '1' is state '0' (probability .71) and the second most likely event following '1' is to persist in that state (probability .12). One can continue in this way comparing the similarities and differences that occur for each subsequent row.

Special attention should be drawn to one additional pattern that emerges in comparing the remaining rows. For the female infant (rows '2' and '3') we note that the most likely outcome following state '2' is a persistence of this state, while the second most likely outcome is the occurrence of state '3'. For row '3' the most likely outcome is a persistence of the same state '3' and next most often is a return to state '2'. What this indicates is that a frequent change of vocalization events for this infant and her mother is for them to vocalize together, followed by mother vocalizing alone, followed by infant vocalizing along with mother, and so on. The table allows us to identify what can be called a frequent "subcycle" of two events (state '2' and state '3') that tend to alternate with each other. A somewhat different subcycle emerges when we examine the male infant and his mother for state '2' and state '3'. This same pattern only holds for the male infant if we ignore the largest entries in rows '2' and '3' and observe the second and third largest. This would suggest that for the female infant and her mother this subcycle tends not to be interrupted while for the male it tends to be interrupted by the '0' state.

Alternative Actions if a Poor Fit Is Obtained. For the infant data so far simulated, the above estimation and simulation procedures work quite well. However, this may not always be the case for every mother-infant interaction. The question then arises, if a particular mother-infant pair seem to yield data which give a very poor fit in terms of matching statistics with the Markov model, should one abandon the possibility that a Markov model can be found for this pair or are there alternative procedures that should be attempted.[3]

There are at least three alternative actions that can be considered:

(a) It may be that the previous definitions of vocalization states are still appropriate for this hypothetical mother-infant pair but that their behavior is dependent upon which overall activity they are engaged in at certain times of the day. For example, it may be that this mother and infant vocalize only in such special situations as "changing the baby's diaper" or "washing the baby" and they tend not to interact when the infant is "lying in its crib or playpen." These situational variants occur for every mother-infant pair that we have studied, but these changes do not appear to have interfered to any great extent with the adequacy of the Markovian model to fit the overall vocalization sequences. Should a poor fit arise though one might consider constructing Markov models for each of the special situations. Naturally one would have to collect a very large number of observations in order to get relatively stable transitional probability estimates for each of these situations. This would be quite time consuming but it does open up the interesting possibility that situational variants may ultimately prove valuable in gaining further insight into the nature of mother-infant interactions. That is, observing the fluctuations in the schedule that occur from day to day (and across different mother-infant pairs), one might eventually be led to study the chaining of these situations themselves (where the situations would then be defined as the relevant states of the mother-infant system). At this juncture in our data collection, though, it is merely an interesting speculation.

(b) A second alternative would be the following: Suppose that the definition of the vocalization states are still adequate but that the unit of time is inappropriate. It is not difficult to imagine that with a 10 sec. interval several states may actually occur in rapid succession with brief pauses between their successive appearance. While this difficulty did not occur to any great extent in the data collected thus far, one would certainly grant that shorter time intervals would eliminate much of this uncertainty in deciding which <u>one</u> of the states actually occurred in the unit interval.

(c) The third option, should a poor fit obtain initially, is to consider the possibility that the vocalization state that occurred on trial n-2 may be significantly influencing which state can occur on trial n. One can take into account this dependency of the current state (on trial n) on both trial n-2 and trial n-1 by redefining what is meant by a state of the mother-infant vocalization system. This redefinition should be considered only as a last resort because it greatly increases the number of transitional probabilities that have to be estimated in order to simulate the vocalization data.

Since there are six possible vocalization events on trial n-2 and the same six possible events for trial n-1, this means that there are a total of 36 possible pairs of events that can affect the current event on trial n. The 36 events are listed below as pairs of events with the first member of the ordered pair referring to the outcome of the vocalization system on trial n-2 and with the second member of the ordered pair referring to the outcome on trial n-1. The 36 events are: 0,0; 0,1; 0,2; 0,2i; 0,3; 0,3i; 1,0; 1,1; 1,2; 1,2i; 1,3; 1,3i; 2,0; 2,1; 2,2; 2,2i; 2,3; 2,3i; 2i,0; 2i,1; 2i,2; 2i,2i; 2i,3; 2i,3i; 3,0; 3,1; 3,2; 3,2i; 3,3; 3,3i; 3i,0; 3i,1; 3i,2; 3i,2i; 3i,3; and 3i,3i. A new transition matrix with 36 rows (with each row labelled by one of the ordered pairs listed immediately above and in that order) and with 36 columns (with each column similarly labelled and in that order) can be constructed.

Consider now how to tally the data sequences using this more complex system of 36 states. The sequence is the same as that used earlier: 3, 0, 1, 3, 1, 2,... and so on. The first and second entries 3,0 are used to locate the row of the new transition matrix and the second and third entries 0,1 are used to locate the appropriate column. Place a tally then in the row labeled 3,0 and the column labeled 0,1. Next use the second and third entries 0,1 to locate the row and the third and fourth entries 1,3 to locate the column. So place a tally in the 0,1 row and the 1,3 column, etc. Notice that the raw data is still scored just as before; the only difference now is that we use a longer string of the data to make each tally. After this larger matrix has been estimated, conditional probabilities for each row would be calculated as before and the initial probability vector would be determined by dividing the sum of the tallies in each row by the total number of tallies in the matrix. Again, it needs emphasis that this more complex matrix is only of interest should the simpler approach fail to achieve a good fit with the data.

Footnotes

[1]This research was supported by the National Institute of Child Health and Human Development, under Research Grant 1 P01 HD01762.

[2]Data presented earlier by Lewis (this volume) indicate that it is possible that other forms of behavior are associated with either infant or maternal vocalization. For the benefit of simplicity only vocalization data are presented.

[3]Although it sounds contradictory at this point, we shall show later in this paper how one might include the effects of trial n-2 as well as n-1 on the current trial should a poor fit between data and the initial Markov model occur. This inclusion of the effects of trial n-2 is done by redefining what is meant by a state of the vocalization system.

[4]Several additional distinctions can be drawn about the different types of Markov processes that can occur. For example Markov state transitions which are independent of the trial number are said to be Markov chains.

[5]It is possible to reduce the number of conditional probabilities that need to be estimated in order to carry out this simulation. One such simplification suggested by examining the full transitional matrices for several mother-infant pairs is that the two most important entries (the two largest entries) in each row tend to be the '0' column entry and the main diagonal entry. One simplifying assumption then might be that if one uses the data to estimate just these two row entries, the remaining row entries will distribute the remaining probability equally among themselves. Another example of how one might simplify these matrices is to consider letting the probability of each entry in the 3i column be equal to the product of the proportions for category '1' and category '2i' that occur in each row.

Table 1

The Markov Transition Matrix and Initial Probability Vector[a] for Six Vocalization States[b]

State on Trial n

State on Trial n-1	0	1	2i	2	3i	3
0	.42	.09	.13	.22	.02	.12
1	.22	.46	.00	.08	.02	.22
2i	.18	.04	.51	.12	.05	.10
2	.05	.01	.05	.71	.01	.17
3i	.27	.13	.20	.07	.07	.26
3	.05	.06	.01	.33	.02	.53

[a]The initial probability vector for the six states 0, 1, 2i, 2, 3i, and 3, respectively, was .13, .07, .09, .44, .02, and .25.

[b]State '0' means neither mother nor infant vocalized; state '1' means only the infant vocalized; state '2' means only the mother vocalized; state '2i' means the mother vocalized to another person; state '3' means both mother and infant vocalized in same time period; and state '3i' means mother vocalized to another and the infant vocalized in the same period. These data apply to the female infant.

Table 2

Predicted and Observed Frequencies of Consecutive Ten Second Period Vocalizations, Mother Vocalizations, and Simultaneous Vocalization of Mother and Infant[a]

Event	Duration of Vocalization	Observed	Predicted (Simulated)
Infant			
Alone	10 sec.	15	16
"	20 "	5	9
"	30 "	4	2
"	40 "	2	1
"	50 "	1	0
"	60 "	0	1
Mother			
Alone	10 "	44	35
"	20 "	13	17
"	30 "	6	10
"	40 "	6	12
"	50 "	6	4
"	60 or more sec.	17	22
Mother & Infant Both in Same Time			
Period	10 sec.	45	48
"	20 "	22	15
"	30 "	6	10
"	40 "	9	4
"	50 "	1	3
"	60 or more sec.	3	3

[a]This table applies to the female infant.

Table 3

The Markov Transition Matrix and Initial Probability Vector[a]

for Six Vocalization States[b]

State on Trial n-1	State on Trial n: 0	1	2i	2	3i	3
0	.77	.05	.07	.08	.01	.02
1	.71	.12	.07	.05	.00	.05
2i	.43	.05	.42	.06	.03	.01
2	.57	.06	.01	.28	.01	.07
3i	.56	.11	.11	.11	.00	.11
3	.41	.11	.07	.15	.00	.26

[a]The initial probability vector for the six states 0, 1, 2i, 2, 3i, and 3 was .70, .06, .10, .09, .01, and .04, respectively.

[b]State '0' means neither mother nor infant vocalized; state '1' means only the infant vocalized; state '2' means only the mother vocalized; state '2i' means the mother vocalized to another person; state '3' means both mother and infant vocalized in same time period; and state '3i' means mother vocalized to another and the infant vocalized in the same period. These data apply to the male infant.

Table 4

Predicted and Observed Frequencies of Consecutive Ten Second Period Vocalizations, Mother Vocalizations, and Simultaneous Vocalization of Mother and Infant[a]

Event	Duration of Vocalization	Observed	Predicted (Simulated)
Infant Alone	10 sec.	34	28
"	20 "	4	6
"	30 "	0	0
Mother Alone	10 "	38	37
"	20 "	9	8
"	30 "	3	2
"	40 "	1	0
"	50 "	0	0
Mother & Infant Both in Same Time Period	10 "	14	11
"	20 "	4	4
"	30 "	0	2
"	40 "	0	0

[a]This table applies to the male infant.