AN ANALYSIS OF SOME CONDITIONS FOR REPRESENTING N STATE MARKOV PROCESSES AS GENERAL ALL OR NONE MODELS

THEODORE E. STEINER

UNIVERSITY OF ILLINOIS

AND

JAMES G. GREENO

UNIVERSITY OF MICHIGAN

Recently Markov learning models with two unidentifiable presolution success states, an error state, and an absorbing learned state, have been suggested to handle certain aspects of data better than the three state Markov models of the General All or None model type. In attempting to interpret psychologically, and evaluate statistically the adequacy of various classes of Markov models, a knowledge of the relationship between the classes of models would be helpful. This paper considers some aspects of the relationship between the class of General All or None models and the class of Stationary Absorbing Markov models with N error states, and M presolution success states.

Many recent Markov models can be considered as special cases of the General All or None model [Greeno & Steiner, 1964], which is represented by the following matrix P and arbitrary initial vector V:

$$(1) \quad P = \begin{array}{c} \\ \mathcal{A} \\ \mathcal{S}' \\ \mathcal{E} \end{array} \begin{array}{ccc} \mathcal{A} & \mathcal{S}' & \mathcal{E} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-w & w \\ u & (1-u)(1-v) & (1-u)V \end{bmatrix} \end{array}, \quad V = (P(\mathcal{A}_1), P(\mathcal{S}_1), P(\mathcal{E}_1))$$

If we let the data take on two values on each trial, say success and error, then the data sequences consist of a finite string of successes (S), and errors (E) and end with an infinite string of successes. The process is in state $\mathcal{E}$ on each trial in which an E occurs, it is in state $\mathcal{S}'$ on each S prior to the last E, and it enters state $\mathcal{A}$ on the trial after the last E. Note that the term "Absorbing Markov Chain" will be used to mean absorption in a success state. Polson

and Greeno [1965] have shown that this class of models need not have the "stationarity" property, that is $P(E_n \mid E_n \cup S_n)$ is not constant independent of n, contrary to what was previously thought. In addition, recently several four-state models which contain two preabsorption indistinguishable "success" states have been suggested [Atkinson & Crothers, 1964] for experiments with the same type of binary data as described above. With the loss of stationarity as a criterion for all or none models, and with the advent of four-state models, several questions arise. First, are there some simple properties that characterize all models that satisfy (1) for success, error sequences? Second, if a N-state model does not imply (1), what statistical properties of data will distinguish the model from (1), and how can we test the null hypothesis that (1) gives a satisfactory fit to the data? Third, what restrictions on the N-state model are necessary for it to satisfy equation 1?

The present discussion is restricted to consideration of the identifiability of transition parameters. In applications, questions about initial probabilities depend at least partly on procedures that vary from one experiment to another. An analysis general enough to cover most procedures would involve a considerable amount of calculation that would not bear on the question of distinguishing N-state processes from (1), following the first error. Therefore, we initially consider the question of whether a model satisfies (1) regarding data starting with the first error. This is a natural way to narrow the problem for the all-or-none model. Errors are recurrent Markov events, which means that everything that occurs before the first error can be ignored in calculating probabilities of things that happen later. We say that a model T, with parameters numerically specified, is a member of the General All-or-None Model Type if the model satisfies (1) following the first error. A model type T will belong to the General All-or-None model type if for every probability measure $P(T)$ generated by a set of parameter values for T, the associated probability distribution conditional on the first error having occurred satisfies equation (1). Of course, some model types may satisfy equation (1) after the first error only with certain restricted sets of parameter values. Because the concern in the present paper is with the statistical properties of model types satisfying equation (1), a model type that satisfies equation (1) after the first error only with parameter restrictions will still be considered a member of the General All-or-None model type so long as the parameter restrictions are imposed.

We will say that a state is observable if for every trial n in any complete response sequence in the outcome space of the experiment, we can say whether or not the state occurred on trial n. (The terminology differs from an earlier article [Greeno & Steiner, 1964] where the term "identifiable" was used to refer to these states. We now use the term "identifiable" to refer to parameters, to be consistent with standard usage in statistics and econometrics [Greeno & Steiner, 1967].)

Analysis of General All-or-None Models for Binary Data

Recurrence Analysis

In the following discussion, reference will often be made to indistinguishable states. The results will, of course, be applicable to the case in which the states are distinguishable, but for purposes of analysis, are considered indistinguishable.

The following notation will be used. Suppose a Markov Chain has a set of error states $\{E_1, \cdots, E_N\}$, a class of success states $\{S_1, \cdots, S_M\}$ and a class of learned states $\{L_1, \cdots, L_R\}$. We will let the events

$$\mathcal{E}_n = \bigcup_{i=1}^{N} E_{i_n},$$

$$\mathcal{S}_n = \bigcup_{i=1}^{M} S_{i_n},$$

and

$$\mathcal{L}_n = \bigcup_{i=1}^{R} L_{i_n},$$

so that the event $\mathcal{E}$, $\mathcal{S}$, or $\mathcal{L}$ will occur on any trial on which one of E_i, S_i, or L_i respectively occur. For any sequence, let n' denote the trial number of the last $\mathcal{E}$. Define f_{J_n} and h_{J_n} as follows:

$$f_{J_n} = P(\mathcal{E}_{n+J} \cap \mathcal{S}_{n+J-1} \cap \cdots \cap \mathcal{S}_{n+1} \mid \mathcal{E}_n), \qquad J \geq 1, \quad n \geq 1$$

and

$$h_{J_n} = P(\mathcal{E}_{n+J} \cap \mathcal{E}_{n+J-1} \cap \cdots \cap \mathcal{E}_{n+1} \mid \mathcal{E}_n), \qquad J \geq 1, \quad n \geq 1.$$

If $\mathcal{E}$ is a recurrent event, as it will be in the cases considered in this paper, the trial subscript, n, can be dropped. Similarly, define $f_J(i)$ and $h_J(i)$ for $i \varepsilon \{1, \cdots, N\}$ as

$$f_{J_n}(i) = P(\mathcal{E}_{n+J} \cap \mathcal{S}_{n+J-1} \cap \cdots \cap \mathcal{S}_{n+1} \mid E_{i_n}), \qquad J \geq 1, \quad n \geq 1$$

and

$$h_{J_n}(i) = P(\mathcal{E}_{n+J} \cap \mathcal{E}_{n+J-1} \cap \cdots \cap \mathcal{E}_{n+1} \mid E_{i_n}), \qquad J \geq 1, \quad n \geq 1.$$

The trial subscript, n, will be dropped because the $\{E_i\}$ are assumed to be Markov states, and therefore are recurrent events. For $J, K \geq 1, i \varepsilon \{1, \cdots, N,\}$ let

$$f(i) = \sum_{J=1}^{\infty} f_J(i),$$

$$s_K(i) = \sum_{J=1}^{K} f_J(i).$$

The concept of lumpability [Burke & Rosenblatt, 1958; Kemeny & Snell, 1960] will be used extensively.

Definition 1. We will say that a Markov chain with states $X(1)$, $\cdots$, $X(m)$ $\cdots$ is lumpable with respect to a partition of the states $Y(1)$, $\cdots$, $Y(n)$, $\cdots$ if the transition probabilities between the $Y(n)$ are constant and independent of the trial number, and if the transition probabilities do not depend upon the choice of a start vector.

Kemeny et al. [1960] have proved that a necessary and sufficient condition for lumpability with respect to the $\{Y(i)\}$ partition is that for every pair $Y(i)$ and $Y(J)$, the transition probability $P_{X(m),Y(i)}$ has the same value for every $X(m)$ in $Y(i)$.

Theorem 1. An Absorbing Markov chain T with stationary transition probabilities, indistinguishable error states $\{E_1, \cdots E_N\}$, indistinguishable success states $\{S_1, \cdots, S_M\}$, and indistinguishable learned states $\{L_1, \cdots, L_r\}$ is a member of the General All or None Model type (G. A. N.) for all initial vectors such that $P_1(\mathcal{E}) = 1$ if and only if for $i \varepsilon \{1, \cdots, N\}$,

(2) $f_J(i) = P(\mathcal{E}_{n+J} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid E_{in})$

$$= \begin{cases} \alpha\beta^{J-1}, & J > 1, \quad 0 \leq \alpha, \beta \leq 1 \\ \gamma, & J = 1, \quad 0 \leq \gamma \leq 1, \end{cases}$$

(3) $h_J(i) = P(\mathcal{E}_{n+J} \cap \cdots \cap \mathcal{E}_{n+1} \mid \mathcal{E}_{in}) = \gamma^J, \quad J \geq 1, \quad 0 \leq \gamma \leq 1.$

Proof. Let us expand the states of T in the following way. Let the $\{E_i\}$ states remain the same. Let T be in $S'_{k,n+k}(i)$ when E_i occurs on trial n, and S occurs on trials $n+1, \cdots, n+k$ and there will be a future occurrence of $\mathcal{E}$. Thus $\{S'_{k,n+k}(i)\}$, for $n, k \geq 1, i \varepsilon \{1, \cdots, N\}$ is the set of all occurrences of S prior to the last occurrence of $\mathcal{E}$. Define T to be in State $\mathcal{C}$ on trial n if $\mathcal{L}$ occurs on n, or if $\mathcal{E}$ does not occur for $n' \geq n$.

Let us say T is in S' on trial n if one of the $\{S'_{k,n+k}(i)\}$ occurs on trial n. The transition probabilities can be written as follows for $n \geq 1, i \varepsilon \{1, \cdots, N\}$,

(4) $P(S'_{k+L,n+k+1}(i) \mid S'_{k,n+k}(J))$

$$= \begin{cases} 0, & i \neq j \quad \text{or} \quad L \neq 1 \\ \dfrac{f(i) - s_{k+1}(i)}{f(i) - s_k(i)}, & i = J, \quad L = 1, n, \quad k \geq 1, \end{cases}$$

(5) $P(\mathcal{E}_{n+k+1} \mid S'_{k,n+k}(i)) = \dfrac{f_k(i)}{f(i) - s_k(i)},$

(6) $P(\mathcal{E}_{n+1} \mid E_{i_n}) = h_1(i),$

(7) $P(S'_{k,n+1}(i) \mid E_{J_n}) = \begin{cases} 0, & i \neq j, \quad k \neq 1 \\ f(i) - f_1(i), & i = J, \quad k = 1, \end{cases}$

(8) $P(\mathcal{C}_{n+1} \mid E_{i_n}) = 1 - f(i).$

If T is assumed to belong to the G. A. N., then (1) is assumed to hold on the states $\mathcal{Q}$, $\mathcal{E}$, $\mathcal{S}'$. Thus

$$P(\mathcal{E}_{n+1} \mid \mathcal{S}'_n) = w$$

$$P(\mathcal{Q}_{n+1} \mid \mathcal{E}_n) = u = 1 - f$$

$$P(\mathcal{E}_{n+1} \mid \mathcal{E}_n) = (1 - u)v = f_1$$

If we let

(9)

$$\alpha = (1 - u)(1 - v) \frac{w}{1 - w}$$

$$\beta = 1 - w$$

$$\gamma = (1 - u)v$$

and substitute in (2), then it can be shown that

$$f_J = \begin{cases} (1 - u)(1 - v)w(1 - w)^{J-2}, & J \geq 2 \\ (1 - u)v, & J = 1, \end{cases}$$

$$h_J = ((1 - u)v)^J, \qquad\qquad\qquad J \geq 1.$$

The trial subscript, n, can be dropped on f_{J_n} because (1) implies that $\mathcal{E}$ is a recurrent event. But it still must be shown that (2) and (3) also hold for all $i \, \varepsilon \, \{1, \cdots, N\}$. Assume there exists a $k \, \varepsilon \, \{1, \cdots, N\}$ such that

$$f_J \neq f_J(k).$$

Consider the initial vector with

$$P_1(E_k) = 1.$$

Clearly then, T would not satisfy equation (1). The same argument holds for the $\{h_J(i)\}$. Therefore, (2) is implied, with α, β, and γ defined as in this proof.

To show the second half of the theorem we assume (2) and (3) hold, and then demonstrate that $\mathcal{Q}$, $\mathcal{S}'$, and $\mathcal{E}$ are Markov states with the appropriate transition probabilities.

Let F be any event on $(\mathcal{Q}, \mathcal{S}', \mathcal{E})$ that depends only on trials after trial n. Let G be any event on $(\mathcal{Q}, \mathcal{S}', \mathcal{E})$ that depends only on trials prior to n such that,

$$P(\mathcal{E}_n \cap G) \neq 0.$$

Consider

$$P(F \mid \mathcal{E}_n \cap G) = \sum_{i=1}^{N} \frac{P(F \mid E_{i_n} \cap G)}{P(\mathcal{E}_n \cap G)} P(E_{i_n} \cap G).$$

Now, since $\{E_i\}$ are all Markov states, and equations (2) and (3) are assumed,

$$P(F \mid E_{i_n} \cap G) = P(F \mid E_{i_n}) = P(F \mid \mathcal{E}_n), \quad i \, \varepsilon \, \{1, \cdots, N\}, \, n \geq 1.$$

Therefore, we can write

$$P(F \mid \mathcal{E}_n \cap G) = P(F \mid \mathcal{E}_n), \qquad n \geq 1.$$

Hence $\mathcal{E}$ is a Markov state. To show that $S'_{k,n+k}(i)$ is a Markov state for $k \geq 1$, $n \geq 1$, and $i \, \varepsilon \, \{1, \cdots, N\}$ assume

$$P(S'_{k,n+k}(i) \cap G) \neq 0.$$

Then, by the definition of $S'_{k,n+k}(i)$ and the fact that the $\{E_i\}$ are Markov states,

$$P(F \mid S'_{k,n+k}(i) \cap G) = P(F \mid \mathcal{S}_{n+k} \cap \cdots \cap \mathcal{S}_{n+1} \cap E_{i_n} \cap G \cap n' > n)$$

$$= P(F \mid \mathcal{S}_{n+k} \cap \cdots \cap \mathcal{S}_{n+1} \cap E_{i_n} \cap n' > n) = P(F \mid S'_{k,n+k}(i))$$

Thus, $S_{k,n+k}(i)$ is a Markov state for $i \, \varepsilon \, \{1, \cdots, N\}$, $k, n \geq 1$. Clearly $\mathcal{C}$ is a Markov state since it is absorbing.

We will now consider the transition probabilities. By definition, for $k, n \geq 1$, $i \, \varepsilon \, \{1, \cdots, N\}$,

(10) $$P(\mathcal{C}_{n+k+1} \mid S'_{k,n+k}(i)) = 0,$$

(11) $$P(\mathcal{C}_{n+1} \mid \mathcal{C}_n) = 1,$$

(12) $$P(\mathcal{E}_{n+1} \mid \mathcal{C}_n) = P(S'_{k,n+k}(i) \mid \mathcal{C}_n) = 0.$$

By substituting (2) and (3) into (4), (5), (6), (7), and (8), one obtains for $i \, \varepsilon \, \{1, \cdots, N\}$, $n, k \geq 1$,

(13) $$P(S'_{k+1,n+k+1}(i) \mid S'_{k,n+k}(i)) = \beta,$$

(14) $$P(\mathcal{E}_{n+k+1} \mid S'_{k,n+k}(i)) = 1 - \beta,$$

(15) $$P(\mathcal{E}_{n+1} \mid \mathcal{E}_n) = \gamma,$$

(16) $$P(S'_{1,n+1}(i) \mid \mathcal{E}_n) = \frac{\alpha\beta}{1 - \beta},$$

(17) $$P(A_{n+1} \mid \mathcal{E}_n) = 1 - \gamma - \frac{\alpha\beta}{1 - \beta}.$$

From (13), (14), and (16), it can be seen that the $\{S'_{k,n+k}(i)\}$ are lumpable to a single Markov state, S'. By the above equations, the transition matrix in (18) directly follows.

$$(18) \qquad P = \begin{array}{c} \\ \mathcal{C} \\ S' \\ \mathcal{E} \end{array} \begin{array}{ccc} \mathcal{C} & S' & \mathcal{E} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & \beta & 1-\beta \\ 1-\gamma-\dfrac{\alpha\beta}{1-\beta} & \dfrac{\alpha\beta}{1-\beta} & \gamma \end{bmatrix} \end{array}.$$

If we let

$$\alpha = (1-u)(1-v)\frac{w}{1-w},$$

$$\beta = (1-w),$$

and

$$\gamma = (1-u)v,$$

(18) is seen to be identical to (1). This completes the proof.

An example of a chain satisfying Theorem 1 is given in (19);

$$(19) \qquad \begin{array}{c} L \\ S_2 \\ S_1 \\ E_1 \\ E_2 \end{array} \begin{array}{ccccc} L & S_2 & S_1 & E_1 & E_2 \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 1/4 & 0 & 1/2 & 1/4 & 0 \\ \dfrac{1+\gamma}{4} & 3/4(1-\gamma) & 0 & \gamma/2 & 0 \\ \gamma/2 & \dfrac{1-\gamma}{2} & \dfrac{1-\gamma}{2} & 0 & \gamma/2 \end{bmatrix} \end{array}$$

with the initial vector $(0, 0, 0, P, 1-P)$.
It can easily be seen that

$$(20) \qquad f_J(1) = f_J(2) = \begin{cases} 3/8(1-\gamma)(1/2)^{J-2}, & J > 1 \\ \gamma/2, & J = 1, \end{cases}$$

$$(21) \qquad h_J(1) = h_J(2) = (\gamma/2)^J, \qquad J \geq 1.$$

In Lemma 1, a condition equivalent to (2) will be given. Let n' denote the trial of last occurrence of $\mathcal{E}$ for a given sequence.

Lemma 1. For an absorbing Markov chain T with error states $\{E_1, \cdots, E_N\}$, success states $\{S_1, \cdots, S_M\}$, and learned states $\{L_1, \cdots, L_R\}$

$$(22) \quad f_J(i) = P(\mathcal{E}_{n+J} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid E_{in}) = \begin{cases} \alpha\beta^{J-1}, & J > 1 \\ \gamma, & J = 1 \end{cases}$$

if and only if for $i \varepsilon \{1, \cdots, N\}$

(23) $g_J(i) = P(\mathcal{E}_{n+J} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid E_{in} \cap n' > n)$

$$= \begin{cases} \dfrac{\alpha\beta^{J-1}}{\gamma + \dfrac{\alpha\beta}{1-\beta}}, & J > 1 \\[6mm] \dfrac{\gamma}{\gamma + \dfrac{\alpha\beta}{1-\beta}}, & J = 1 \end{cases},$$

(24) $$P(n' > n \mid E_{in}) = \gamma + \frac{\alpha\beta}{1-\beta}.$$

Proof. The proof follows directly from a consideration of conditional probabilities.

By Theorem 1 and Lemma 1, the following Theorem is proven.

Theorem 2. An Absorbing Markov Chain T with indistinguishable error states $\{E_1, \cdots, E_N\}$, indistinguishable success states $\{S_1, \cdots, S_M\}$ and indistinguishable learned states $\{L_1, \cdots, L_R\}$ is a member of the General All or None model type for all initial vectors such that $P_1(\mathcal{E}) = 1$ if and only if for $i \varepsilon \{1, \cdots, N\}$, $0 \le \alpha', \beta, \gamma' \le 1$,

(25) $g_J(i) = P(\mathcal{E}_{n+J} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid E_{in} \cap n' > n)$

$$= \begin{cases} \alpha'\beta^{J-1}, & J > 1 \\ \gamma', & J = 1 \end{cases},$$

(26) $$P(n' > n \mid E_{in}) = \gamma',$$

where

$$\alpha' = \frac{\alpha}{\gamma + \dfrac{\alpha\beta}{1-\beta}},$$

$$\gamma' = \frac{\gamma}{\gamma + \dfrac{\alpha\beta}{1-\beta}},$$

and

(27) $$h_J(i) = P(\mathcal{E}_{n+J} \cap \cdots \cap \mathcal{E}_{n+1} \mid E_{in}) = \gamma', \quad J > 0.$$

Theorem 2 is interesting for the following reasons. If an Absorbing Markov chain with error states $\{E_1, \cdots, E_N\}$ satisfies equation (26), then

(28) $$1 - f(i) = P(n' = n \mid E_{in}) = 1 - \gamma', \quad i \varepsilon \{1, \cdots, N\},$$

since $n' < n$ is impossible if E_i occurred on trial n. Thus,

(29)
$$P(n' = n \mid \mathcal{E}_n) = P\left(n' = n \mid \bigcup_{i=1}^{N} E_{in}\right) = \frac{\sum P(n' = n \mid E_{in})P(E_{in})}{\sum P(E_{in})}$$

$$= (1 - \gamma') \frac{\sum_i P(E_{in})}{\sum_i P(E_{in})} = 1 - \gamma'.$$

Equation (29), and thus (26), can be shown to imply the almost geometric distribution of total errors usually associated with All or None models. Theorem 2 shows that the geometric distribution of total errors is not a strong enough condition to insure membership in the G. A. N.

In the light of Theorem 2, several useful properties result from writing the G. A. N. model in the form of equation (1). The following are (26) and (27) written with the parameters of (1).

(30) $\quad g_J = P(\mathcal{E}_{n+J+1} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid \mathcal{E}_n \cap n' > n)$

$$= \begin{cases} (1 - v)w(1 - w)^{J-1}, & J \geq 1 \\ v, & J = 0. \end{cases}$$

(31) $\quad P(n' > n \mid \mathcal{E}_n) = 1 - u.$

(32) $\quad P(\mathcal{E}_{n+J} \cap \cdots \cap \mathcal{E}_{n+1} \mid \mathcal{E}_n) = [(1 - u)v]^J, \quad J \geq 1.$

The maximum likelihood estimators of u, v, and w, are known [Greeno & Steiner, 1964]. So, if (1), and thus (30), (31), and (32), are assumed to hold, a direct and separate test can be made of the conditions of Theorem 2 by the use of (30), (31), and (32).

In summary, Theorems 1 and 2 provide two sets of equivalent necessary and sufficient conditions for a Markov model with stationary transition probabilities to belong to the G. A. N. model type following the first error. The conditions of Theorem 2 can be written so as to provide separate tests of the necessary and sufficient conditions. Therefore it appears that Theorem 2 might be more useful in a data test. However, in the next section, it will be seen that Theorem 1 is very useful for theoretical analysis.

*Matrix Analysis**

Once Theorem 1 is obtained, a natural question to ask is whether there exist restrictions on transition matrices that are equivalent to the conditions of theorem 1. The following definition of a property which may obtain between classes of states will be useful:

Definition 2. Let $H = \{H_1, \cdots, H_M\}$ and $K = \{K_1, \cdots, K_N\}$ be two classes of Markov states which constitute a subset of the states of some Markov chain. If H and K are disjoint classes, let t_n be a random variable that takes on the number of trials, after trial n, to first passage to K for each sequence that is in H on Trial n, and remains in H until passage to K. If there is a β independent of n and $\mathbf{P}_n(H)$, such that for all probability vectors $\langle P(H_{1_n}), \cdots, P(H_{M_n}) \rangle$ for Trial n,

(33) $P(\tau_n = J \mid H_n) = \alpha'_n \beta^J$ $n \geq 1, \quad J \geq 0, \quad 0 \leq \alpha'_n, \quad \beta \leq 1,$

then we will say that H is unitary relative to entrance into K. If $H = K$, then define τ_n to be a random variable that takes on the number of trials in H after trial n before exit, for each sequence in H on Trial n. Then H is unitary relative to entrance into H, if equation (33) holds.

It is clear that if $M = 1$, that is if H as only one member, then H will be unitary with respect to entrance into any class, because H will be a Markov state, and all Markov states satisfy (33). Also, if the states of H satisfy the Kemeny and Snell [1960] conditions for lumpability, then H will be unitary relative to entrance into any class of states. However, Theorem 3 will provide weaker conditions, not equivalent to the Kemeny et al. conditions, under which H will be unitary.

The following notation will be needed in this section. Let $\mathbf{P}(K \mid H)$ be a vector such that its ith component, $P(K \mid H_i)$, denotes the probability of a one-step transition from H_i to any state in K. Let $\mathbf{P}^r$ denote $\mathbf{P}$ transpose. Also, the notation $[0]$ will be used to denote the zero vector of dimension appropriate to the context.

Theorem 3: Let $H = \{H_1, \cdots, H_M\}$ and $K = \{K_1, \cdots, K_N\}$, N, $M \geq 1$, be two classes of Markov states. Let C be the matrix of stationary transition probabilities from states of H to states of H. H is unitary with respect to entrance into K if and only if there exists a β such that

(34) $(C - \beta I_{M \times M}) \mathbf{P}^r(K \mid H) = [0], \qquad 0 \leq \beta \leq 1.$

Proof. Assume (34) holds. Let $\mathbf{P}_n(H) = \langle P(H_{1_n}), \cdots, P(H_{M_n}) \rangle$ be any probability vector for H, on trial n. Now

$$P(\tau_n = J) = \mathbf{P}_n(H) C^J \mathbf{P}^r(K \mid H), \qquad J \geq 1$$

$$P(\tau_n = 0) = \mathbf{P}_n(H) \mathbf{P}^r(K \mid H).$$

By (34),

(35) $C\mathbf{P}^r(K \mid H) = \beta I \mathbf{P}^r(K \mid H) = \mathbf{P}^r(K \mid H).$

It can be simply shown by induction that (35) implies

$$C^k \mathbf{P}^r(K \mid H) = \beta^k \mathbf{P}^r(K \mid H), \qquad k \geq 1.$$

Thus, we have

$$P(\tau_n = J \mid H_n) = \frac{\mathbf{P}_n(H)\beta^J}{P_n(H)} \mathbf{P}^r(K \mid H) = \beta^J \frac{\mathbf{P}_n(H)\mathbf{P}^r(K \mid H)}{P_n(H)}, \qquad J \geq 1.$$

Let

$$\alpha'_n = \frac{\mathbf{P}_n(H)\mathbf{P}^r(K \mid H)}{P_n(H)}.$$

Then

$$P(\tau_n = J \mid H_n) = \alpha'_n \beta^J, \qquad J \geq 0.$$

Thus H is unitary relative to entrance into K.

To prove the second half of the theorem, assume (33) holds. Now, as was shown in the first half of the proof,

$$(36) \qquad P(\tau_n = J \mid H_n) = \mathbf{P}_n \frac{(H)}{P(H_n)} C^J \mathbf{P}^r(K \mid H), \qquad J \geq 1,$$

$$(37) \qquad P(\tau_n = 0 \mid H_n) = \frac{\mathbf{P}_n(H)}{P(H_n)} \mathbf{P}^r(K \mid H).$$

By (33), (36), and (37), we have

$$(38)$$

$$\frac{\mathbf{P}_n(H)}{P(H_n)} C^J \mathbf{P}^r(K \mid H) = \alpha'_n \beta^J, \qquad n \geq 1, \quad J \geq 1, \quad 0 \leq \alpha'_n, \quad \beta \leq 1,$$

$$\frac{\mathbf{P}_n(H)}{P(H_n)} \mathbf{P}^r(K \mid H) = \alpha'_n.$$

Equation (38) holds for all vectors $\mathbf{P}_n(H)$, with β constant. Consider vectors which start the process out in a single state of H with a probability of 1.

$$\mathbf{P}^r(K \mid H) = (\alpha'_1, \cdots, \alpha'_M)^r$$

$$(1, 0, \cdots, 0)C^J \mathbf{P}^r(K \mid H) = \alpha'_1 \beta^J$$

$$(39) \qquad (0, 1, 0, \cdots, 0)C^J \mathbf{P}^r(K \mid H) = \alpha'_2 \beta^J$$

$$\vdots$$

$$(0, \cdots, 0, 1)C^J \mathbf{P}^r(K \mid H) = \alpha'_M \beta^J.$$

Thus we have shown that

$$(40) \qquad I_{M \times M} C^J \mathbf{P}^r(K \mid H) = C^J \mathbf{P}^r(K \mid H) = \begin{bmatrix} \alpha'_1 & \beta^J \\ \vdots & \\ \alpha'_M & \beta^J \end{bmatrix}$$

By (39)

(41)
$$\beta I_{M \times M} \mathbf{P}^r(K \mid H) = \begin{bmatrix} \alpha_1' & \beta \\ \vdots & \\ \alpha_M' & \beta \end{bmatrix}.$$

Subtracting (41) from (40) and setting

$$J = 1,$$

we have

(42) $CP^r(K \mid H) - \beta I_{M \times M}\mathbf{P}^r(K \mid H) = (C - \beta I_{M \times M})\mathbf{P}^r(K \mid H) = [0].$

Clearly $0 \le \beta \le 1$ by assumption in (33). Thus, the second half of the proof is complete.

   Theorem 3 shows that H is unitary with respect to K if and only if $\mathbf{P}(K \mid H)$ is an eigenvector of C corresponding to an eigenvalue β that can be a probability. The intuitive meaning of the above, and (34) is that the effects of the matrix C operation on the vector $\mathbf{P}(K \mid H)$ can be represented as multiplication by a single constant β.

   The following example shows that the conditions of Theorem 3, and thus equation (34), are not equivalent to lumpability of the states of H. The values of the transition probabilities for E_1, E_2 are only restricted by the usual probability considerations in the example.

   Example 1. Let $H = \{S_1, S_2\}$; $K = \{E_1, E_2\}$.

(43)

	L	S_2	S_1	E_2	E_1
L	1	0	0	0	0
S_2	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	0
S_1	$\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$
E_2	a	b	c	d	e
E_1	f	g	h	j	k

   Equation (43) satisfies Theorem 3 with $\beta = \frac{1}{2}$, but S_1 and S_2 do not meet the Kemeny et al. [1960] conditions for lumpability.

   In order to apply the new definition and Theorem 3 to the problem of finding restrictions on matrices equivalent to the conditions of Theorem 1, some additional notation is necessary. Let

$$\mathbf{f}_J = (f_J(1), \cdots, f_J(N))$$

and

$$\mathbf{P}(\mathcal{E} \mid S) = \langle P(\mathcal{E} \mid S_1), \cdots, P(\mathcal{E} \mid S_M) \rangle,$$

where $P(\mathcal{E} \mid S_J)$ is the probability that an error state occurs on trial $n + 1$

given S_J on trial n. Some symbols are necessary to represent certain submatrices of the stationary one step transition matrix. Let G be the submatrix of transition probability from the $\{E_i\}$ to the $\{S_J\}$, C be the submatrix of transition probabilities from $\{S_J\}$ to $\{S_J\}$, and D be the submatrix of transition probabilities from $\{E_i\}$ to $\{E_i\}$.

The definition of unitary, definition 2, was formulated with respect to all possible probability vectors for the states of H. For the part of this section that deals with restricted vectors, the following definition will be necessary:

Definition 3. Let $H = \{H_1, \cdots, H_N\}$, and $K = \{K_1, \cdots, K_M\}$, for $N, M \geq 1$, be two classes of Markov states which constitute a subset of the states of some Markov chain. If H and K are disjoint classes, let τ_n be a random variable defined as in definition 1. If there is a β such that for all probability vectors for H_n in V, where V is a subset of the set of all possible probability vectors for H on trial n, (33) is satisfied, then H is unitary relative to entrance into K over V. If H equals K, define τ_n as in definition 2 for this case. Then if (33) holds for all probability vectors for H in V, where V is a subset of the set of all possible probability vectors for H on trial n, then H is unitary relative entrance into H over V.

It is easy to produce examples to show that, in general, the condition expressed in (34) is not a necessary condition for definition 2, if V is a proper subset of the set of possible vectors for H_n.

Clearly, if H is unitary with respect to entrance into K, then H will be unitary with respect to entrance into K over V, for all V.

Lemma 2. For an absorbing Markov chain with error states $\{E_1, \cdots, E_N\}$ $N \geq 1$, success states $\{S_1, \cdots, S_M\}$ $M \geq 1$, absorbing learned states $\{L_1, \cdots, L_R\}$, $R \geq 1$, stationary transition probabilities,

$$P_1(\mathcal{E}) = 1,$$

and for $i \, \varepsilon \, \{1, \cdots, N\}$,

(44) $\qquad f_J(i) = P(\mathcal{E}_{n+J} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid E_{in})$

$$= \begin{cases} \alpha\beta^{J-1}, & J > 1, \quad 0 \leq \beta \leq 1 \\ \gamma, & J = 1, \quad 0 \leq \gamma \leq 1, \end{cases}$$

(45) $\qquad h_J(i) = P(\mathcal{E}_{n+J} \cap \cdots \cap \mathcal{E}_{n+1} \mid E_{in}) = \gamma^J, \qquad J \geq 1, \quad 0 \leq \gamma \leq 1$

if and only if

   (A) S is unitary with respect to entrance into $\mathcal{E}$ over set of all vectors such that $P_1(\mathcal{E}) = 1$

   (B) $\mathcal{E}$ is unitary with respect to entrance into $\mathcal{E}$ over set of all vectors such that $P_1(\mathcal{E}) = 1$

(46) $$GP'(\mathcal{E} \mid \mathcal{S}) = (\alpha\beta, \cdots, \alpha\beta)'$$

(47) $$\mathbf{P}(\mathcal{E} \mid \mathcal{E}) = (\gamma, \cdots, \gamma).$$

Proof.

Assume $N \geq 1$ and $M \geq 1$.

Assume (44) and (45). In the proof of Theorem 1, the conditions of (44) and (45) were shown to imply that $\mathcal{E}$ was a Markov state. Since a Markov state is unitary with respect to any class of states, conditions (B) follows immediately.

To show condition (A), first define a partition of the set of all occurrences of $\mathcal{S}$ as follows. Let T be in the state $\mathcal{S}_{K,n+K}$ on trial $n + K$ if

$$(\mathcal{E}_n \cap \mathcal{S}_{n+1} \cap \mathcal{S}_{n+2} \cap \cdots \mathcal{S}_{n+K})$$

occurs.

Clearly, $\{\mathcal{S}_{K,n+K}\}$ for $K \geq 1, n \geq 1$ is a partition of all occurrences of S, because it is assumed that $P(\mathcal{E}_1) = 1$.

Let X_n be any particular history of the process up to trial n. Then, there exists for every occurrence of $\mathcal{S}_n$ a $K \geq 1$ such that for $J \geq 1$,

(48)
$$P(\tau_n = J \mid \mathcal{S}_{K,n}) = P(\mathcal{E}_{n+J+1} \cap \mathcal{S}_{n+J} \cap \cdots \cap \mathcal{S}_{n+1} \mid \mathcal{S}_n \cap X_{n-1})$$

$$= \frac{P(\mathcal{E}_{n+J+1} \cap \mathcal{S}_{n+J} \cap \cdots \cap \mathcal{S}_{n-K+1} \cap \mathcal{E}_{n-K} \cap X_{n-K-1})}{P(\mathcal{S}_n \cap \cdots \cap \mathcal{S}_{n-K+1} \cap \mathcal{E}_{n-K} \cap X_{n-K-1})}$$

$$= \frac{P(\mathcal{E}_{n+J+1} \cap \mathcal{S}_{n+J} \cap \cdots \cap \mathcal{S}_{n-K+1} \mid \mathcal{E}_{n-K})}{P(\mathcal{S}_n \cap \cdots \cap \mathcal{S}_{n-K+1} \mid \mathcal{E}_{n-K})}.$$

The fact that $\mathcal{E}$ is Markov is needed to derive equation (48). Using the f_J notation, we have obtained

$$P(\tau_n = J \mid \mathcal{S}_{K,n}) = \frac{f_{K+J+1}}{P(\mathcal{S}_n \cap \cdots \cap \mathcal{S}_{n-K+1} \mid \mathcal{E}_{n-K})} = \frac{f_{K+J+1}}{\alpha_{n,k}} = \left(\frac{\alpha\beta^K}{\alpha_{n,K}}\right)\beta^J$$

where

$$\alpha_{n,K} = P(\mathcal{S}_n \cap \cdots \cap \mathcal{S}_{n-K+1} \mid \mathcal{E}_{n-K}), \quad J, n, K \geq 1.$$

The value of $\alpha_{n,K}$ is determined by the chain, and the initial vector. Let

$$\alpha'_{n,K} = \frac{\alpha\beta^K}{\alpha_K}.$$

Now, since for $n > 1$,

$$\mathcal{S}_n = \bigcup_{K=1}^{n-1} \mathcal{S}_{K,n},$$

we have for $J \geq 0, n > 1$,

$$(49) \quad P(\tau_n = J \mid \mathcal{S}_n) = \frac{P(\tau_n = J \cap \mathcal{S}_n)}{P(\mathcal{S}_n)} = \sum_{K=1}^{n-1} \frac{P(\tau_n = J \mid \mathcal{S}_{K,n}) P(\mathcal{S}_{K,n})}{P(\mathcal{S}_n)}$$

$$= \beta^J \sum_{K=1}^{n-1} \alpha'_{n,K} \frac{P(\mathcal{S}_{K,n})}{P(\mathcal{S}_n)} = \alpha''_{n,K} \beta^J,$$

where

$$\alpha''_{n,K} = \sum_{K=1}^{n-1} \alpha'_{n,K} \frac{P(\mathcal{S}_{K,n})}{P(\mathcal{S}_n)}.$$

Because (49) holds for all arbitrary start vectors satisfying the restriction that

$$P(\mathcal{E}_1) = 1,$$

(49) implies that $\mathcal{S}$ is unitary with respect to $\mathcal{E}$ over the set of initial vectors such that

$$P(\mathcal{E}_1) = 1.$$

By (44), setting $J = 2$, we have

$$\mathbf{f}_2^r = G\mathbf{P}^r(E \mid S) = (\alpha\beta, \cdots, \alpha\beta)^r.$$

By setting

$$J = 1$$

in (45), (47) follows.

To show the second half of Lemma 2, consider the following. By the definition of $\mathcal{S}$ being unitary over the set of initial vectors such that

$$P(\mathcal{E}_1) = 1$$

and the fact that the probability of $\mathcal{E}$ on trial one is one, we can write

$$(50) \quad \alpha'\beta^J = P(\tau_2 = J \mid \mathcal{S}_2) = P(\tau_2 = J \mid \mathcal{S}_2 \cap \mathcal{E}_1).$$

$$= \frac{f_{J+2,1}}{P(\mathcal{S}_2 \mid \mathcal{E}_1)}, \quad J \geq 0.$$

Thus,

$$f_{J+2,1} = P(\mathcal{S}_2 \mid \mathcal{E}_1)\alpha'\beta^J, \quad J \geq 0.$$

Consider an initial vector such that for some $i \, \varepsilon \, \{1, \cdots, N\}$

$$P(E_i) = 1.$$

Then,

$$(51) \quad f_{J+2,1}(i) = P(\mathcal{S}_2 \mid E_i)\alpha'_i\beta^J = \mathbf{P}_1(\mathcal{E})GC^J\mathbf{P}^r(\mathcal{E} \mid S).$$

Since E_i is Markov,

$$f_J(i) = f_{J_1}(i), \quad \_\ J \geq 1,$$

and

$$P(\mathcal{S}_{n+1} \mid E_{in}) = P(\mathcal{S}_2 \mid E_{i_1}), \quad \_\ n \geq 1.$$

The above argument holds for all $i \ \varepsilon \ \{1, \cdots, N\}$. Thus the following is obtained.

(52) $\mathbf{f}_J^r = IGC^J \mathbf{P}^r(\mathcal{E} \mid \mathcal{S}) = \beta^{J-2}(P(\mathcal{S}_2 \mid E_1)\alpha_1', \cdots, P(\mathcal{S}_2 E_{N_1})\alpha_N')^r, \quad J \geq 2.$

Equation (46) implies that

$$\mathbf{f}_2 = (\alpha\beta, \cdots, \alpha\beta) = (P(\mathcal{S}_2 E_{1_1})\alpha_1', \cdots, P(\mathcal{S}_2 E_{N_1})\alpha_N').$$

Thus,

(53) $\mathbf{f}_J = (\alpha\beta^{J-1}, \cdots, \alpha\beta^{J-1}), \quad J \geq 2.$

Condition D implies that

$$\mathbf{f}_1 = \gamma.$$

A similar argument shows that

$$\mathbf{h}_J^r = D^{J-1} \mathbf{P}^r(\mathcal{E} \mid \mathcal{E}) = (\gamma^{J-1}, \cdots, \gamma^{J-1})^r, \quad J \geq 2,$$

$$h_1 = \gamma.$$

Thus, the second half of Lemma 2 is complete. By Lemma 2 and Theorem 1, Theorem 4 has been proven.

Theorem 4. An absorbing Markov chain T with error states $\{E_1, \cdots, E_N\}$, $N \geq 1$, success states $\{S_1, \cdots, S_M\}$, $M \geq 1$, and absorbing learned states $\{L_1, \cdots, L_R\}$, $R \geq 1$, stationary transition probabilities, and $P(\mathcal{E}_1)=1$ is a member of the G. A. N. if and only if

(A) $\mathcal{S}$ is unitary with respect to entrance into $\mathcal{E}$ over the set of probability vectors for $\mathcal{S}$ such that $P_1(\mathcal{E}) = 1$.

(B) $\mathcal{E}$ is unitary with respect to entrance into $\mathcal{E}$ over the set of probability vectors for $\mathcal{E}$ such that $P_1(\mathcal{E}) = 1$.

(C) $G\mathbf{P}^r(\mathcal{E} \mid \mathcal{S}) = (\alpha\beta, \cdots, \alpha\beta)^r$.

(D) $\mathbf{P}^r(\mathcal{E} \mid \mathcal{E}) = (\gamma, \cdots, \gamma)^r$.

To extend Theorem 4 to a broader class of initial vectors is not difficult. Consider a start vector allowing sequences in $\mathcal{L}$, $\mathcal{S}$, or $\mathcal{E}$ on Trial 1. Clearly no new restrictions are necessary to handle sequences starting in $\mathcal{L}$ or $\mathcal{E}$. Sequences starting in $\mathcal{S}$ that never enter $\mathcal{E}$ also cause no difficulties because they are put into $\mathcal{A}$. Now, consider sequences starting in $\mathcal{S}$ that do enter $\mathcal{E}$. Since $\mathcal{E}$ is a Markov state, once $\mathcal{E}$ is entered, the probability of all future events on the level of A, $\mathcal{E}$, and $\mathcal{S}$ do not depend on what occurred prior to $\mathcal{E}$. Thus the probability of these sequences is just the product of the probability of the initial string of successes terminating in an $\mathcal{E}$, and the probability of the future events given $\mathcal{E}$. All that is necessary then, is to bring the proba-

bilities of the strings of successes that start in S and terminate in ε into line with the strings of successes that are preceded by an error and terminate in ε. It can be shown that the conditions of Theorem 5 are the proper ones.

Theorem 5. An absorbing Markov chain with error states $\{E_1, \cdots, E_N\}$, $N \geq 1$, success states $\{S_1, \cdots, S_M\}$, $M \geq 1$, and absorbing learned states $\{L_1, \cdots, L_R\}$, stationary transition probabilities, is a member of the G. A. N. for all initial vectors if and only if

(A) S is unitary with respect to entrance into ε.
(B) ε is unitary with respect to entrance into ε.

$$G\mathbf{P}'(\varepsilon \mid \mathcal{S}) = (\alpha\beta, \cdots, \alpha\beta)^{\tau}$$
$$\mathbf{P}'(\varepsilon \mid \varepsilon) = (\gamma, \cdots, \gamma)^{\tau}$$
$$G\mathbf{P}'(\varepsilon \mid \mathcal{S}) = \beta\mathbf{P}'(\varepsilon \mid \varepsilon).$$

The major results of this section are the following. Previously, one might have intuitively thought that the only way a large Markov chain with some classes of unobservable states might belong to the G. A. N. was either that the appropriate states would lump by the Kemeny et al. [1960] theorem, or that the chain could be expanded into new states that did satisfy the Kemeny et al. [1960] theorem. Theorems 4 and 5 prove that the G. A. N. contains a much larger class of models than one might have previously expected.

It is important to note that, in general, conditions for inclusion in the G. A. N. will depend upon the class of initial vectors over which membership is required. Thus, if in a given experiment, the data is generated by a particular Markov chain, whether or not the data will be fit by a G. A. N. model may depend upon the value of the initial vector.

Application of Matrix Analysis to Four State Models

In this section, some results are presented regarding an unrestricted four-state model, with errors occurring in only one state. A table of some general restrictions sufficient for membership in the G. A. N. is given. Also, some extant models are analyzed from the present perspective.

Let the following matrix P represent a four-state Markov model with one error state (E), two unobservable presolution success states, S_1 and S_2, and a learned state (L),

$$
(54) \qquad P = \begin{array}{c} \\ L \\ S_2 \\ S_1 \\ E \end{array} \begin{array}{c} L \quad\ S_2 \quad\ S_1 \quad\ E \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ P_{2,3} & P_{2,2} & P_{2,1} & P_{2,0} \\ P_{1,3} & P_{1,2} & P_{1,1} & P_{1,0} \\ P_{0,3} & P_{0,2} & P_{0,1} & P_{0,0} \end{bmatrix} \end{array}
$$

and some initial vector $\mathbf{V} = (P_0(L), P_0(S_2), P_0(S_1), P_0(E))$.

TABLE 1

Example Restrictions on Equation 54 Making P Satisfy G.A.N. Following First Error for $(P_{2,0}, P_{1,0})$

General Restrictions	$v = w$ Restriction
Case I $P_{1,2} \cdot P_{2,1} \neq 0$	
$\quad P_{1,1} \cdot P_{2,2} = P_{1,2} \cdot P_{2,1}$	$P_{0,0}(P_{1,1} + P_{2,2}) = P_{0,1} \cdot P_{1,1}$
	$\qquad\qquad\qquad\qquad + P_{0,2}P_{2,0}$
$\quad P_{1,1} + P_{2,2} \leq 1$	
$\quad (P_{2,0}, P_{1,0}) = (P_{1,0}P_{2,2}/P_{1,2}, P_{1,0})$	
Case II $P_{1,2} \cdot P_{2,1} = 0$	
$\quad P_{1,2} = P_{2,1} = 0$	$P_{0,0} \cdot P_{2,2} = P_{0,1}P_{1,0} + P_{0,2}P_{2,0}$
$\quad P_{1,1} = P_{2,2}$	

Since there is only a single error state,

$$P(\mathcal{E}_{n+J+1} \cap \cdots , \cap \mathcal{E}_{n+1} \mid \mathcal{E}_n) = P_{0,0}^{J+1} , \qquad J \geq 0$$

and,

$$GP^r(E \mid \mathcal{S}) = P_{0,2}P_{2,0} + P_{0,1}P_{1,0}).$$

By Theorem 5, all that remains to insure that (54) satisfies (1) after the first error is that $\mathcal{S}$ is unitary with respect to $\mathcal{E}$. Table 1 lists several ways in which (34) is satisfied for $P_{2,0}$ and $P_{1,0}$. The entries in Table 1 were obtained by solving the characteristic equation [Murdoch, 1957] of the submatrix C in (54), and obtaining solutions such that $(P_{2,0}, P_{1,0})$ is an eigenvector corresponding to a $0 \leq \beta \leq 1$. This is the general procedure for obtaining constraints on $(P_{2,0}, P_{1,0})$.

$$C = \begin{bmatrix} P_{2,2} & P_{2,1} \\ P_{1,2} & P_{1,1} \end{bmatrix}$$

So, if a model fulfilling the restrictions of (54) fulfills a condition in Table 1, it will be a member of the G. A. N. model type following the first error.

By calculating the recursion for each case in Table 1, column 1, the restrictions necessary for the $v = w$ case of (1) can be obtained. The restrictions for $v = w$ corresponding to the column 1 restrictions are contained in column 2.

In addition to the restrictions in Table 1, Column 1, other specific solutions are obtainable by the same methods, if restrictions are put on $(P_{2,0}, P_{1,0})$.

One unusual case of (54) which satisfies (1) following the first error is

$$P_{1,1} = P_{2,2} = 0,$$

$$P_{1,2} \cdot P_{2,1} \neq 0,$$

$$P_{1,0} = P_{2,0}\sqrt{\frac{P_{1,2}}{P_{2,1}}}.$$

As an example, consider Bernbach's [1965] forgetting model for paired-associate learning. We give a slightly expanded expression of the model. There are five states: L, where the correct response is in permanent memory; H, where the correct response is in temporary memory; I, where an incorrect response is in temporary memory; G, where no response is in memory, but where the subject guesses correctly; and E where no response is in memory, and the subject guesses incorrectly. The transition probabilities are

$$(55) \quad P = \begin{array}{c} \\ L \\ H \\ G \\ E \\ I \end{array} \begin{array}{ccccc} L & H & G & E & I \\ \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ c & (1-c)h & (1-c)(1-h)g & (1-c)(1-h)(1-g) & 0 \\ 0 & h & (1-h)g & (1-h)(1-g) & 0 \\ 0 & (1-e)h & (1-h)g & (1-h)(1-g) & eh \\ 0 & (1-e)h & (1-h)g & (1-h)(1-g) & eh \end{array}\right] \end{array}$$

The distinction between states I and E has implications for a special statistic, the probability of repeating the same error on successive trials. However, in estimating parameters and in the main evaluation of the model, Bernbach considered just the sequences of correct responses and errors. For the data considered in this way, states I and E can be considered as a single state W, where wrong responses occur. Then Bernbach's model is in the form of (54), with state H corresponding to S_2, G to S_1, and W to E. According to Table 1, Bernbach's model is in the G. A. N. model type.[†] We have, in general,

$$P_{1,2}P_{2,1} = h(1 - c)(1 - h)g \neq 0;$$

which is Case I. Then we see that

$$P_{1,1}P_{2,2} - P_{1,2}P_{2,1} = (1 - h)g(1 - c)h - h(1 - c)(1 - h) = 0.$$

The restriction $P_{1,1} + P_{2,2} \leq 1$ is satisfied for most parameter values. However, we have

$$P_{0,0}(P_{1,1} + P_{2,2}) = P_{0,1}P_{1,0} + P_{0,2}P_{2,0}$$

if and only if $e = 0$. Therefore, Bernbach's model usually will not generate

[†] All examples in this section can be shown to satisfy the Case I restriction on $(P_{2,0}, P_{1,0})$.

data agreeing with all-or-none learning with $v = w$; however, it will generate data agreeing with the general all-or-none model of (1).

As a second example, consider one case of a model investigated by Greeno [1967]. There are four states: L, where an item is in permanent memory; H, where the item is in temporary memory; E, where the item is not in memory and the subject gives a wrong response; and G, where the item is not in memory but the subject guesses correctly. The transition probabilities are

$$(56) \quad P = \begin{array}{c} \\ L \\ H \\ G \\ E \end{array} \begin{array}{cccc} L & H & G & E \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ c & (1-c)h & (1-c)(1-h)g & (1-c)(1-h)(1-g) \\ d & (1-d)h & (1-d)(1-h)g & (1-d)(1-h)(1-g) \\ d & (1-d)h & (1-d)(1-h)g & (1-d)(1-h)(1-g) \end{bmatrix} \end{array}$$

The model as given here relates to Bernbach's rather closely. If e in Bernbach's model is zero, and if d in (56) is zero, then the two models are identical. Further, if c and d are equal in (56), we have the model that Atkinson and Crothers [1964] called LS-2.

Applying the present analysis to the model, we can see that the model should imply (1) with $v = w$. Using Table 1, we first see that we are in Case I, since, in general,

$$P_{1,2}P_{2,1} = (1 - d)h(1 - c)(1 - h)g \neq 0.$$

Then within Case I,

$$P_{1,1} + P_{2,2} = (1 - d)(1 - h)g + (1 - c)h > 0$$

and simple algebra shows that,

$$P_{1,1}P_{2,2} = P_{1,2}P_{2,1}$$

The result is consistent with earlier findings; Greeno's [1967] article has a proof that (56) implies (1) with $v = w$.

Finally, consider a more general form of (56),

$$(57) \quad P = \begin{array}{c} \\ L \\ S_2 \\ S_1 \\ E \end{array} \begin{array}{cccc} L & S_2 & S_1 & E \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ e & (1-e)XP & (1-e)(1-X)P & (1-e)(1-P) \\ d & (1-d)Yg & (1-d)(1-Y)g & (1-d)(1-g) \\ c & (1-c)ZS & (1-c)(1-Z)S & (1-c)(1-S) \end{bmatrix} \end{array}$$

One obtains:

$$P_{1,1}P_{2,2} - P_{1,2}P_{2,1}$$

$$= (1 - d)(1 - Y)g(1 - e)XP - (1 - d)Yg(1 - e)(1 - X)P$$

$$= (1 - d)(1 - e)gP[X(1 - Y) - Y(1 - X)]$$

$$= (1 - d)(1 - e)gP[X - Y].$$

If one assumes, as in (56) and the LS-2 Model,

$$0 \neq P_{1,1} + P_{2,2} \leq 0$$

and

$$P_{1,2} \cdot P_{2,1} \neq 0,$$

if $X = Y$, then

$$P_{1,1} \cdot P_{2,2} = P_{1,2} \cdot P_{2,1}.$$

Thus with all the apparent complexity of (57), such a model, following the first error, will imply the G. A. N. model type if $P = g$, as can be seen by applying the Case I restriction for $(P_{2,0}, P_{1,0})$.

Application of Recurrence Analysis

In the previous section, the matrix analysis was applied to an analysis of models. In this section, two applications of the recurrence analysis are made. First, a likelihood ratio test is derived testing a four state theory against a three state theory. Second, the restrictions imposed upon (54) if (1) is satisfied are examined.

As a specific alternative to the three-state all-or-none theory, consider the following four-state theory, which is a generalization of models investigated by Atkinson and Crothers [1964] and Greeno [1967]. The states are L, where the item is learned; H, where the item is in temporary memory; G, where the item is not in memory but the subject guesses correctly; and E, where the item is not in memory and the subject gives an incorrect response. The transition probabilities are

$$(58) \quad P = \begin{array}{c} L \\ H \\ G \\ E \end{array} \begin{bmatrix} \begin{array}{cccc} L & H & G & E \\ 1 & 0 & 0 & 0 \\ c & (1-c)h_1 & (1-c)(1-h_1)g & (1-c)(1-h_1)(1-g) \\ d & (1-d)h_2 & (1-d)(1-h_2)g & (1-d)(1-h_2)(1-g) \\ d & (1-d)h_2 & (1-d)(1-h_2)g & (1-d)(1-h_2)(1-g) \end{array} \end{bmatrix},$$

$$0 < c, d, h_1, h_2, g < 1.$$

In the case where $h_1 = h_2$, Greeno [1967] showed that this model implies G. A. N. with $v = w$.

We can show quite easily, using theorems 3 and 5, that for $h_1 \neq h_2$ the model is not a member of the G. A. N. over the set of all initial vectors. All that must be done is to show that $\langle P(E_{n+1} \mid H_n), P(E_{n+1} \mid G_n) \rangle$ is an eigenvector of

$$
C = \begin{array}{c} \\ H \\ G \end{array} \begin{array}{c} \overset{H}{} \\ \begin{bmatrix} (1 - c)h_1 & (1 - c)(1 - h_1)g \\ (1 - d)h_2 & (1 - d)(1 - h_2)g \end{bmatrix} \end{array}
$$

if and only if

$$
h_1 = h_2 .
$$

The proof requires simple algebra.

However, if we restrict the initial vector such that

$$
P_1(E) = 1,
$$

then we know that the requirement that $\mathrm{P}(\mathcal{E} \mid \mathcal{S})$ be an eigenvector of C satisfies one of the conditions of theorem 4, but may not be a necessary condition. It can be shown that (58) could belong to the G. A. N., with $v = w$, over the set of initial vectors such that

$$
P_1(E) = 1
$$

if and only if

$$
h_1 = h_2 .
$$

Thus, for the two classes of initial vectors under consideration, the equality of h_1 and h_2 is necessary and sufficient for membership in G. A. N., with $v = w$.

We will present some results which permit a statistical test of (1), in the framework of (58). Specifically, the null hypothesis is that in (58),

$$
h_1 = h_2 ;
$$

and the alternative is that (58) holds but that

$$
0 < h_1 , \qquad h_2 < 1.
$$

The test to be developed is a likelihood ratio test, and we deal with identifiable parameters. Atkinson and Crothers [1964] partially solved the problem of identification for (58) by showing that the following three parameters are sufficient and identifiable.

$$
r = (1 - c)h_1 ,
$$

(59)
$$
s = (1 - d)(1 - h_2),
$$

$$
t = (1 - c)(1 - d)(1 - h_1)h_2 .
$$

To use this result in an application, we need the likelihood of data as a function of r, s, and t. The recurrence probabilities give a convenient summary of the probability measure. We will show that

$$f_J = s(1 - g)(a - b)^J$$

(60)
$$+ \frac{(1 - g)}{2b} [t + s^2 g - s(a - b)][(a + b)^J - (a - b)^J],$$

$$f = \sum_{J=0}^{\infty} f_J = (1 - g)\left[\frac{(1 - r)s + t}{(1 - a)^2 - b^2}\right],$$

where

$$a = \frac{r + sg}{2},$$

$$b = \frac{\sqrt{(r + sg)^2 - 4(rs - t)}}{2} \neq 0.$$

To begin we define the quantities

$$X_J = P(H_{n+J} \cap C_{n+J-1} \cap \cdots, \cap C_{n+1} \mid E_n),$$

$$Y_J = P(G_{n+J} \cap C_{n+J-1} \cap \cdots, \cap C_{n+1} \mid E_n).$$

From (58) we have

$$X_{J+1} = (1 - c)h_1 X_J + (1 - d)h_2 Y_J,$$

$$Y_{J+1} = (1 - c)(1 - h_1)g X_J + (1 - d)(1 - h_2)g Y_J,$$

which holds for all $J \geq 0$ if we set $X_0 = 0$ and $Y_0 = 1$. We can translate these recursions into two second-order difference equations.

$$X_{J+2} - (1 - c)h_1 + (1 - d)(1 - h_2)g X_{J+1} + (1 - c)(1 - d)(h_1 - h_2)X_J$$

$$= X_{J+2} - (r + sg)X_{J+1} + (rs - t)X_J = 0;$$

$$Y_{J+2} - (r + sg)Y_{J+1} + (rs - t)Y_J = 0.$$

The solution of a second-order difference equation is described by Goldberg [1958]. In the present case we have

$$X_J = K_1 m_1^J + K_2 m_2^J,$$

$$Y_J = K_3 m_1^J + K_4 m_2^J,$$

where

$$m_1 = \frac{(r + sg) + \sqrt{(r + sg)^2 - 4g(rs - t)}}{2} = a + b,$$

$$m_2 = \frac{(r + sg) - \sqrt{(r + sg)^2 - 4g(rs - t)}}{2} = a - b.$$

If $m_1 = m_2$, a different set of equations apply [Goldberg, 1958]. The constants are determined by the initial conditions.

$$X_0 = 0,$$

$$X_1 = (1 - d)h_2,$$

$$Y_0 = 1,$$

$$Y_1 = (1 - d)(1 - h_2)g.$$

Then we obtain

$$K_2 = -K_1 = \frac{-(1 - d)h_2}{2b},$$

$$K_4 = 1 - K_3 = 1 - \frac{(1 - d)(1 - h_2)g - (a - b)}{2b}.$$

We now obtain the f_J :

$$f_J = X_J(1 - c)(1 - h_1)(1 - g) + Y_J(1 - d)(1 - h_2)(1 - g)$$

$$= \frac{(1 - d)h_2}{2b} [(a + b)^J - (a - b)^J](1 - c)(1 - h_1)(1 - g)$$

$$+ \left\{ \frac{(1 - d)(1 - h_2)g - (a - b)}{2b} [(a + b)^J - (a - b)^J] + (a - b)^J \right\}$$

$$\cdot (1 - d)(1 - h)(1 - g)$$

and substitution of r, s, and t gives (60).

We do not see a way to develop closed-form estimates of r, s, and t from (60). However, estimates would be easy to obtain using iterative search on a computer. The general hypothesis consists of (60). Since errors are recurrent Markov events, each error can be considered as the beginning of an independent observation of the system. We can describe the data after the first error by tabulating the values of n_i, the frequency of success runs of length i, where n_∞ is the number of times that an error is followed by no more errors. Then the likelihood function is

$$L = (1 - f)^{n_\infty} f_0^{n_0} f_1^{n_1} f_2^{n_2}, \cdots,$$

which will be a finite product for any set of experimental results.

The restricted hypothesis is (60) with $h_1 = h_2$. This is obtained in the identifiable parameters by the restriction

$$t = rs.$$

Note that this specializes the recurrence probabilities so that $a = b$; hence,

$$f_J = \frac{(1 - g)}{2b} (rs + s^2 g)(a + b)^J = (1 - g)s(r + sg)^J.$$

Then r and s could be estimated easily by direct calculation. However, an iterative search program could be written to maximize the likelihood under the restriction using the general estimation but with the search restricted to the $t = rs$ plane.

Now we will turn to our second topic of this section. What constraints are imposed on (54) by requiring (1) to be satisfied also. Suppose a set of data satisfies (1), following the first error, with u, v, and w estimated by using the maximum likelihood estimation procedure [Greeno, et al., 1964]. One then obtains

$$f_J = (1 - \hat{u})(1 - \hat{v})(1 - \hat{w})^{J-1}\hat{w}, \qquad J > 1,$$

where $\hat{u}$, $\hat{v}$, and $\hat{w}$ are maximum likelihood estimates. In terms of Theorem 1, one obtains

$$\hat{\alpha} = (1 - \hat{u})(1 - \hat{v})\hat{w},$$

$$\hat{\beta} = 1 - \hat{w}.$$

The restrictions imposed on (54) have to be examined in two cases.

Case 1.   Assume,

$$P_{1,2} \cdot P_{2,1} \neq 0.$$

Let

(61) $$X_J = P(S_{2n+J} \cap S_{n+J-1} \cap \cdots , \cap S_{n+1} \mid E_n)$$

(62) $$Y_J = P(S_{1n+J} \cap S_{n+J-1} \cap \cdots \cap S_{n+1} \mid E_n)$$

It can be easily shown that

(63) $$X_{J+2} = aX_{J+1} + bX_J = 0$$

(64) $$Y_{J+2} = aY_{J+1} + bY_J = 0,$$

where

$$a = P_{1,1} + P_{2,2}$$

$$b = P_{1,1}P_{2,2} - P_{1,2}P_{2,1} .$$

Now, using (63) and (64), one can show that

(65) $$f_{J+2} = X_{J+1}P_{2,0} + Y_{J+1}P_{1,0} = af_{J+1} - bf_J .$$

Since it is assumed that (1) is satisfied, the $\{f_J\}$ is almost geometric, and hence,

(66) $$f_{J+2} = \beta^J f_2 , \qquad J \geq 0.$$

Combining (65) and (66), one obtains

(67) $$\beta^2 = a\beta - b = \beta(P_{1,1} + P_{2,2}) - (P_{1,1}P_{2,2} - P_{1,2}P_{2,1}),$$

if

$$\beta f_2 \neq 0.$$

Since $\hat{\beta} = 1 - \hat{w}$ is fixed by the estimation procedure, (67) provides a restriction on (54). In addition, the following restrictions on (54) also apply.

(68) $\qquad f_2 = P_{0,1}P_{1,0} + P_{0,2}P_{2,0} = (1 - \hat{u})(1 - \hat{v})\hat{w}$

(69) $\qquad f_1 = P_{0,0} = (1 - \hat{u})\hat{v}.$

Case 2.

$$P_{1,2} \cdot P_{2,1} = 0$$

Starting with (61) and (62), it can be shown that

(70) $\qquad f_J = K_1 Z_1^{J-1} + K_2 Z_2^{J-1} , \qquad J > 1,$

where K_1 , K_2 , Z_1 , Z_2 depend upon the particular instance of Case 2 being examined. [Goldberg, 1961]. Case 2 splits up into many special cases, but the following analysis holds for all cases.

In order to use the same general technique that was used in Case 1, let us ask if there is a pair (a, b) such that

(71) $\qquad f_{J+2} = a f_{J+1} - b f_J , \qquad J \geq 1$

where f_J is not identically zero for all J.

Using (70) and (71), one obtains

(72) $\underline{K_1 Z_1^{J-1}(Z_1^2 - aZ_1 + b) + K_2 Z_2^{J-1}(Z_2^2 - aZ_2 + b) = 0,} \qquad J \geq 1.$

If

$$Z_1 \neq Z_2 ,$$

then for (72) to hold and for the f_J distribution not to be identically equal to 0 for all $J > 1$,

(73) $\qquad Z_1^2 - aZ_1 + b = 0$

(74) $\qquad Z_2^2 - aZ_2 + b = 0.$

Equations (73) and (74) provide 2 equations in a and b. Assuming Z_1 , Z_2 both not zero, one obtains

(75) $\qquad a = Z_1 + Z_2$

(76) $\qquad b = -Z_1 Z_2 .$

If one of Z_1 , Z_2 is zero, but both are not zero, one obtains the solution that

$$a = Z \neq 0$$

$$b = 0.$$

If, however, Z_1 and Z_2 are equal, then if f_J is not equal to zero for all $J > 1$,

$$Z^2 - aZ + b = 0.$$

Thus, there always exist a pair (a, b) for Case 2 such that (71) holds. Again, assuming that the data satisfy (1), (66) can be also assumed. Combining (66) and (71), (67) is shown to hold for Case 2. As in Case 1, the additional restrictions in (68) and (69) also clearly apply to Case 2.

This analysis provides a method to examine various cases of (54) in order to analyze the restrictions imposed upon (54) assuming it satisfies (1) following the first error. An extension of the same type of analysis to runs of successes prior to the first error is also possible.

REFERENCES

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. *Journal of Mathematical Psychology,* 1964, 1, 285–315.

Bernbach, H. A. A forgetting model for paired-associate learning. *Journal of Mathematical Psychology,* 1965, 2, 128–144.

Burke, C. J., & Rosenblatt, M. A Markovian function of a Markov chain. *Annals of Mathematical Statistics,* 1958, 29, 1112–1122.

Goldberg, S. *Introduction to difference equations.* New York: Wiley & Sons, Inc., 1961.

Greeno, J. G. Paired-associate learning with short term retention: mathematical analysis and data regarding identification of parameters. *Journal of Mathematical Psychology,* 1967, 4, 430–472.

Greeno, J. G. Identifiability and statistical properties of two-stage learning with no successes in the initial stage. *Psychometrika,* 1968, 33, 173–215.

Greeno, J. G., & Steiner, T. E. Markovian processes with identifiable states: General considerations and application to all or none learning. *Psychometrika,* 1964, 29, 309–333.

Greeno, J. G., & Steiner, T. E. Comments on "Markovian processes with identifiable states: General considerations and application to all or none learning." *Psychometrika,* 1968, 33, 169–172.

Jolley, L. B. W. *Summation of series* (2nd ed.). New York: Dover, 1961.

Kemeny, J. G., & J. L. Snell, *Finite Markov Chains.* Princeton: D. Van Nostrand Co., 1960.

Murdoch, D. C. *Linear algebra for undergraduates.* New York: John Wiley & Sons, 1957.

Polson, P. G., & Greeno, J. G. Nonstationary performance before all or none learning. Paper read at Midwestern Psychological Association, 1965.

Suppes, P., & Ginsberg, Rose. A fundamental property of all or none models. *Psychological Review,* 1963, 70, 139–161.