JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 7, 167-175 (1963)

Reduced Forms for Stochastic Sequential Machines*

J. W. CARLYLE

Engineering Department, University of California,
Los Angeles, California

Submitted by Lotfi Zadeh

I. INTRODUCTION

A *stochastic sequential machine* $M = (X, Y, S, P)$ is defined through the specification of finite sets X, Y, and S (whose elements are called *input symbols*, *output symbols*, and *states*, respectively), together with a conditional probability function

$$P(y; s' \mid s; x) : x \in X; y \in Y; s, s' \in S. \tag{1.1}$$

The model M is supposed to represent an idealized physical system with a finite number of distinct internal configurations (states), such that (i) if input x is applied when the state is s, then $P(y; s' \mid s; x)$ is the joint probability that the observed response is y and the new state is s', and (ii) if the machine is initially in state s_1 and inputs $x_1, x_2, \cdots, x_n$ are applied successively, then the output sequence $y_1, y_2, \cdots, y_n$ has the probability distribution

$$p_{s_1}(y_1y_2 \cdots y_n \mid x_1x_2 \cdots x_n) = \sum_{s_k \in S, k \geqslant 2} \prod_{k=1}^{n} P(y_k; s_{k+1} \mid s_k; x_k). \tag{1.2}$$

If s, $t \in S$ and for every positive integer n,

$$p_s(y_1y_2 \cdots y_n \mid x_1x_2 \cdots x_n) = p_t(y_1y_2 \cdots y_n \mid x_1x_2 \cdots x_n) \tag{1.3}$$

for all $x_k \in X$, $y_k \in Y$, $k = 1, 2, \cdots, n$, then states s and t are said to be *equivalent* (i.e., s and t are indistinguishable as internal initial conditions, since they give rise to the same "externally observable behavior" or "input-output relation" $p_s(\cdot \mid \cdot)$). In Section II (Theorem 1) it is established that a sufficient condition for the equivalence of any two states s and t is that (1.3) hold for

* This research was conducted in part while the author was a National Science Foundation Cooperative Graduate Fellow, University of California; it was also supported in part by the Office of Naval Research under Contract Nonr-222(53).

$n = c - 1$, where c is the number of elements in S. This is a stochastic generalization of Moore's basic result [1] for deterministic finite-state machines (in the present context, M is *deterministic* if its probability function P assumes only the values 0 and 1; conventional operational descriptions can evidently be recast in this form). The proof of Theorem 1 is suggested by the work of Blackwell and Koopmans [2] and Gilbert [3] on functions of finite-state stationary Markov chains. The output sequence of a stochastic sequential machine can be expressed as a function of a finite-state *controlled* Markov chain (for which the state transition probabilities at any time n depend upon the input symbol applied at time n). Such a model has been employed [4] for the discrete noisy communication channel; the equivalent formulation given by (1.1) and (1.2) was introduced by Shannon [5] for this purpose. Thus the results obtained here can be interpreted either as stochastic generalizations of corresponding facts in deterministic automata theory, or as structural properties of finite-state communication channels.

If some of the states of a stochastic sequential machine are equivalent, then a portion of the machine's specifications must be redundant; in Section III it is shown that, as in the deterministic case, such redundancies can be eliminated to yield a simplified version of the machine, called a reduced form. Reduced forms for stochastic sequential machines may not be unique (in contrast with the deterministic case); there is, however, a computational procedure for finding all of the reduced forms of a given machine.

II. A Sufficient Condition for Equivalence

Let $M = (X, Y, S, P)$ be a machine with c states; for notational convenience we identify S with the set of integers $1, 2, \cdots, c$. For each pair (x, y) of input and output symbols, let $M(y \mid x)$ be the matrix whose ij element is

$$m_{ij}(y \mid x) = P(y; j \mid i; x).$$

Then $\{M(y \mid x) : x \in X, y \in Y\}$ is a family of $c \times c$ matrices with nonnegative elements such that for each x, the *state transition matrix*

$$M(x) = \sum_y M(y \mid x)$$

is a Markov matrix; any family $\{M(y \mid x)\}$ of matrices with these properties determines a c-state machine. Thus, as an alternate machine notation, we write $M = \{M(y \mid x)\}$; the input and output sets X and Y are assumed common to all machines under consideration and need not be mentioned explicitly. The letters u and v are used to denote finite sequences of input and output symbols respectively, including the empty sequence ϕ having zero length; uu' means the sequence u followed by u', and $|u|$ is the length of u, so that, for example, $|yv| = 1 + |v|$. When both u- and v-sequences appear in the same context (with identical superscripts or subscripts if needed), it is assumed that $|u| = |v|$.

If $\{M(y \mid x)\}$ is a c-state machine, then the $c \times c$ matrices $M(v \mid u)$ are defined by

$$M(v \mid u) = M(y_1 \mid x_1)\, M(y_2 \mid x_2) \cdots M(y_n \mid x_n) \tag{2.1}$$

when $u = x_1 x_2 \cdots x_n$ and $v = y_1 y_2 \cdots y_n$, while $M(\phi \mid \phi)$ is the identity matrix. If e is the c-component column vector with all components equal to 1, then (1.2) shows that

$$h(v \mid u) = M(v \mid u)\, e \tag{2.2}$$

where $h(v \mid u)$ is a c-component column vector whose i-th component is $p_i(v \mid u)$. From (2.1) and (2.2), we obtain a method for constructing h-vectors recursively:

$$h(vv' \mid uu') = M(v \mid u)\, h(v' \mid u'). \tag{2.3}$$

If $\pi = (\pi_1, \pi_2, \cdots, \pi_c)$ is any probability distribution on the states of M, i.e., any *initial distribution for* M, then the number

$$p_\pi(v \mid u) = \pi h(v \mid u) \tag{2.4}$$

is the probability that v is the response when u is applied starting with initial condition π. If k is a nonnegative integer, two initial distributions π and λ are said to be *k-equivalent* (with respect to M) provided that $p_\pi(v \mid u) = p_\lambda(v \mid u)$ for all u and v of length k; π and λ are *equivalent* if they are k-equivalent for all k (otherwise, *distinguishable*). When π and λ are degenerate distributions, this definition evidently agrees with the previously introduced concept of equivalence of states.

THEOREM 1. *If M is a machine with c states, and if π and λ are any two initial distributions for M, then $(c-1)$-equivalence of π and λ is a sufficient condition for their equivalence.*

PROOF. If z is any c-component column vector, let $\psi(z) = (\pi - \lambda)\, z$; in particular, then, using (2.4),

$$\psi(h(v \mid u)) = p_\pi(v \mid u) - p_\lambda(v \mid u),$$

so that π and λ are k-equivalent if and only if ψ vanishes on the linear subspace (of c-dimensional space) spanned by the vectors $h(v \mid u)$ for all u and v

of length k. Let this subspace be $L_k(M)$; the theorem follows if it can be established that for sufficiently large k (i.e., $k \geq c - 1$), all spaces $L_k(M)$ are identical. First, observe that for any k, $L_k \subset L_{k+1}$, since

$$h(v \mid u) = \sum_y h(vy \mid ux).$$

On the other hand, if $L_k = L_{k+1}$ then $L_{k+1} = L_{k+2}$, since L_{k+2} is spanned by the vectors $h(yv \mid xu)$ for all x and y and all u and v of length $k + 1$, and for such x, y, u, v,

$$\begin{aligned} h(yv \mid xu) &= M(y \mid x)\, h(v \mid u) \\ &= M(y \mid x)\,(\text{vector in } L_{k+1}) = M(y \mid x)\,(\text{vector in } L_k) \\ &= M(y \mid x) \left\{ \begin{matrix} \text{linear combination of the} \\ \text{vectors } h(v' \mid u') : \mid u' \mid = \mid v' \mid = k \end{matrix} \right\} \\ &= \{\text{linear combination of } h(yv' \mid xu')\} \in L_{k+1}\,. \end{aligned}$$

Thus there is an integer $J = J(M)$ such that $L_k = L_J$ for all $k > J$ while the dimension of L_{k+1} is strictly greater than the dimension of L_k for all $k < J$. In particular,

$$(\dim L_0) + J \leq \dim L_J \, (\leq c).$$

But $\dim L_0 = \dim \{\text{space spanned by } e\} = 1$ for any machine, so that $J \leq c - 1$, which completes the proof.

When two or more machines are being considered simultaneously, it is convenient to attach identifying superscripts to the quantities $p_\pi(v \mid u)$ and $h(v \mid u)$; e.g., (2.2) might be written $h^M(v \mid u) = M(v \mid u)\, e$. Suppose that π and λ are initial distributions for machines M and N respectively (M and N need not have the same number of states); if $p_\pi^M(v \mid u) = p_\lambda^N(v \mid u)$ for all u and v of length k, we say that *initial distribution π for M is k-equivalent to initial distribution λ for N*, or that *(M, π) and (N, λ) are k-equivalent systems.* If k-equivalence holds for all k, then (M, π) and (N, λ) are *equivalent*, symbolized by $(M, \pi) \sim (N, \lambda)$. If π is concentrated on a single state s, we may write $(M, s) \sim (N, \lambda)$; i.e., states and degenerate distributions are interchangeable in the notation.

THEOREM 2. *If M has c states, N has d states, and π and λ are initial distributions for M and N respectively, then $(c + d - 1)$-equivalence of (M, π) and (N, λ) is a sufficient condition for their equivalence.*

PROOF. Let $M + N$ be the *sum* machine defined by

$$(M + N)(y \mid x) = \left(\begin{array}{c|c} M(y \mid x) & 0 \\ \hline 0 & N(y \mid x) \end{array} \right).$$

Apply Theorem 1 to the machine $M + N$ and the initial distributions

$$\pi' = (\pi_1, \pi_2, \cdots, \pi_c, 0, 0, \cdots, 0),$$

$$\lambda' = (0, 0, \cdots, 0, \lambda_1, \lambda_2, \cdots, \lambda_d).$$

The sufficient conditions given in Theorems 1 and 2 are the best obtainable depending only on the number of states, since simple deterministic machines are easily constructed for which the bounds $c - 1$ and $c + d - 1$ are actually attained [1].

Theorem 2 with $M = N$ was established for functions of Markov chains (i.e., autonomous stochastic sequential machines) by Gilbert [3], and previously (with a bound $2c^2 + 1$ rather than $2c - 1$) by Blackwell and Koopmans [2]. It is the latter method of proof which has been modified to yield the present Theorems 1 and 2. A portion of the procedure used by Moore [1] for deterministic machines, involving a hierarchy of partitions of the state set with respect to k-equivalence, $k = 1, 2, \cdots$, can be extended to the stochastic case [6], provided that "state set" is replaced by "set of all initial distributions."

It is an immediate consequence of Theorem 2 (with $M = N$) that if the structure of M is not given, but $p_\pi^M(v \mid u)$ is specified for all sequences of length $2c - 1$, then $p_\pi^M(v \mid u)$ is uniquely determined for all sequences of arbitrary length. For functions of Markov chains, an explicit rule for the construction of the function p_π^M from its values for arguments of length $2c - 1$ has been obtained by Gilbert [3]; this result can be extended to apply to any stochastic sequential machine [6].

III. Reduction

If to each state of machine M there corresponds an equivalent state of machine N and to each state of N there corresponds an equivalent state of M, then M and N are said to be *state-equivalent machines*. Among the machines which are state-equivalent to a given machine M, those having the smallest number of states are called *reduced forms of M*. A machine for which any two states are distinguishable (i.e., there are no equivalent states) is said to be *in reduced form*. The terminology is consistent, since the reduced forms of any machine M are precisely those machines which are state-equivalent to M and in reduced form; this follows from Theorem 3 below, and from the transitivity of state equivalence.

THEOREM 3. *Let M be a c-state machine with at least one pair of equivalent states. Then there exist $(c - 1)$-state machines which are state-equivalent to M.*

In particular, if s and t are equivalent states of M, let $N(y \mid x)$ be the matrix obtained from $M(y \mid x)$ by deleting row t and column t and replacing column s with the sum of columns s and t; then $N = \{N(y \mid x)\}$ is a $(c - 1)$-state machine which is state-equivalent to M.

PROOF. For notational convenience, let the states of M be renumbered in such a way that $s = c - 1$ and $t = c$. Then

$$n_{ij}(y \mid x) = m_{ij}(y \mid x), \qquad i = 1, 2, \cdots, c - 1, \qquad j = 1, 2, \cdots, c - 2,$$
$$n_{i,c-1}(y \mid x) = m_{i,c-1}(y \mid x) + m_{ic}(y \mid x), \qquad i = 1, 2, \cdots, c - 1.$$

It is clear that $\{N(y \mid x)\}$ is a machine. We wish to show that for all u, v,

$$p_i^N(v \mid u) = p_i^M(v \mid u), \qquad i = 1, 2, \cdots, c - 1, \tag{3.1}$$

and

$$p_{c-1}^N(v \mid u) = p_c^M(v \mid u). \tag{3.2}$$

Observe that (3.2) follows from (3.1) with $i = c - 1$, since states c and $c - 1$ of machine M are assumed to be equivalent. The asserted equalities are evident (by construction of N) for $|u| = |v| = 1$, and if they hold for all sequences u and v of length k, then for such sequences and for any x and y,

$$p_i^M(yv \mid xu) = \sum_{j=1}^{c} m_{ij}(y \mid x)\, p_j^M(v \mid u)$$
$$= \sum_{j=1}^{c-1} n_{ij}(y \mid x)\, p_j^N(v \mid u) = p_i^N(yv \mid xu)$$

for $i = 1, 2, \cdots, c - 1$, so that (3.1) holds for sequences of length $k + 1$, which completes the proof by induction.

The rule for construction of machine N in Theorem 3, called *the merging of equivalent states*, is generalization of the familiar reduction procedure employed in the theory of deterministic sequential machines. The process of merging equivalent states can be applied repeatedly to a given machine M until a reduced form of M is obtained. If M is deterministic, it is well known that its reduced form is unique (apart from permutations of the state set); however if M is stochastic, it may possess a continuum of distinct reduced forms, as shown by Theorem 4 and the subsequent discussion and example.

THEOREM 4. *Let M be a machine having c states; let $L = L_J(M)$ be the associated linear space introduced in Theorem 1, and let H be a $c \times (\dim L)$*

matrix whose columns form a basis for L. Let $\{N(y \mid x)\}$ be any set of $c \times c$ matrices satisfying

$$N(y \mid x)\, H = M(y \mid x)\, H, \tag{3.3a}$$

$$n_{ij}(y \mid x) \geq 0, \qquad i, j = 1, 2, \cdots, c. \tag{3.3b}$$

Then $N = \{N(y \mid x)\}$ is a c-state machine which is state-equivalent to M; in fact,

$$(N, s) \sim (M, s), \qquad s = 1, 2, \cdots, c. \tag{3.4}$$

Conversely, if a c-state machine satisfies (3.4), then it must also satisfy (3.3).

PROOF. The converse follows immediately from (2.3) and the fact that (3.4) can be rewritten in the form

$$h^N(v \mid u) = h^M(v \mid u) \qquad \text{for all} \qquad u, v. \tag{3.5}$$

For the direct assertion, first observe that if $\{N(y \mid x)\}$ is a set of matrices satisfying (3.3), then in particular (since $e \in L$)

$$N(y \mid x)\, e = M(y \mid x)\, e. \tag{3.6}$$

Summing over all y (for any fixed x), we obtain $N(x)\, e = M(x)\, e = e$, verifying (with (3.3b)) that $N(x)$ is a Markov matrix; thus $\{N(y \mid x)\}$ is a machine. Assertion (3.4) (or (3.5)) follows by induction on $|\, v \,|$; it has been established for $|\, v \,| = 1$ in (3.6), and if it holds for all sequences u and v of length k, then for such sequences and for any x and y,

$$\begin{aligned} h^N(yv \mid xu) &= N(y \mid x)\, h^N(v \mid u) = N(y \mid x)\, h^M(v \mid u) \\ &= M(y \mid x)\, h^M(v \mid u) = h^M(yv \mid xu), \end{aligned}$$

using (2.3) and (3.3a).

If M is a c-state machine with a input symbols and b output symbols, the set $\{N(y \mid x)\}$ of solutions of Eq. (3.3a) depends upon $abc\,(c - \dim L)$ unrestricted real parameters. According to Theorem 4, if one restricts the ranges of the parameters in such a way that (3.3b) is always satisfied, this parametric family of sets $\{N(y \mid x)\}$ of matrices coincides (apart from permutations of the states) with the family of all c-state machines which are state-equivalent to M. Applying this result to a reduced form of M, say M' (which might be obtained by repeated application of Theorem 3 to M), we see that the family of all reduced forms of M may depend upon as many as $abc'\,(c' - \dim L)$ parameters, where c' is the number of states of M', i.e., the number of equivalence classes of states of M (it is apparent from their definitions that $\dim L(M)$ cannot exceed c' and that $\dim L(M') = \dim L(M)$). Thus M has a unique reduced form if $\dim L = c'$, while M may possess a continuum of distinct reduced forms (as in the example below) when $\dim L < c'$. (It should be noted that the condition $\dim L < c'$ can also occur when the reduced form is unique; a simple deterministic machine enjoying these properties is given in [6].)

Example. Let $X = Y = \{0, 1\}$, $S = \{1, 2, 3\}$, and

$$M(0 \mid 0) = \begin{pmatrix} .2 & .2 & .2 \\ .2 & .2 & .2 \\ .2 & .2 & .2 \end{pmatrix}, \qquad M(1 \mid 0) = \begin{pmatrix} .4 & 0 & 0 \\ .1 & .3 & 0 \\ .2 & .1 & .1 \end{pmatrix},$$

$$M(0 \mid 1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & .6 \\ .6 & .1 & .1 \end{pmatrix}, \qquad M(1 \mid 1) = \begin{pmatrix} 0 & 0 & 0 \\ .4 & 0 & 0 \\ .2 & 0 & 0 \end{pmatrix}.$$

It is evident that M is already in reduced form; i.e., there are no equivalent states, since (for example) the three components of $h(1 \mid 1) = M(1 \mid 1)\, e$ are distinct. Calculation of the vectors $h(v \mid u)$ for $| v | = 1$ and 2 shows that $J = 1$ (i.e., $L_1 = L$) and that $\dim L = 2\,(< 3 = c = c')$. A convenient choice for the matrix H of Theorem 4 is

$$H = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

When Eqs. (3.3a) are solved with constraints (3.3b), we obtain the following 7-parameter family of all distinct forms $\{N(y \mid x)\}$ for M:

$$N(0 \mid 0) = \begin{pmatrix} A & A & .6 - 2A \\ B & B & .6 - 2B \\ C & C & .6 - 2C \end{pmatrix}, \qquad N(1 \mid 0) = \begin{pmatrix} .4 & 0 & 0 \\ D & D + .2 & .2 - 2D \\ E & E - .1 & .5 - 2E \end{pmatrix},$$

$$N(0 \mid 1) = \begin{pmatrix} 1 & 0 & 0 \\ F & F & .6 - 2F \\ G & G - .5 & 1.3 - 2G \end{pmatrix},$$

$$N(1 \mid 1) = \begin{pmatrix} 0 & 0 & 0 \\ .4 & 0 & 0 \\ .2 & 0 & 0 \end{pmatrix},$$

$$0 \leq A, B, C, F \leq .3, \qquad 0 \leq D \leq .1, \qquad .1 \leq E \leq .25, \qquad .5 \leq G \leq .65.$$

Selection of the minimum value for each parameter yields a machine with particularly simple structure. On the other hand, if maximum values are chosen for all parameters, it becomes apparent that the machine M, although in reduced form, can be "reduced" further in the following sense: the two-state machine M^*, defined by

$$M^*(0 \mid 0) = \begin{pmatrix} .3 & .3 \\ .3 & .3 \end{pmatrix}, \qquad M^*(1 \mid 0) = \begin{pmatrix} .4 & 0 \\ .1 & .3 \end{pmatrix},$$

$$M^*(0 \mid 1) = \begin{pmatrix} 1 & 0 \\ .3 & .3 \end{pmatrix}, \qquad M^*(1 \mid 1) = \begin{pmatrix} 0 & 0 \\ .4 & 0 \end{pmatrix},$$

is such that $(M, 1) \sim (M^*, 1)$, $(M, 2) \sim (M^*, 2)$, and $(M, 3) \sim (M^*, \pi)$, where $\pi = (.5, .5)$. (For purposes of example, the machine M was constructed with this property in mind.) This is not an admissible reduction of M from the "mechanical" viewpoint; we have tacitly assumed that machines commence operation with some fixed initial state. However, the concept of a totally reduced form (for which states of the original machine may correspond to distributions on the new minimal state set) may be germane to the unsolved problem of identification of those finite-state systems whose output sequences have the same probability structure as a given controlled stochastic process [2, 3, 6].

Acknowledgment

The author wishes to express his appreciation to Professor A. J. Thomasian who proposed the problem studied here and made many helpful comments.

References

1. Moore, E. F. Gedanken-Experiments on Sequential Machines. In "Automata Studies." Annals of Mathematics Studies No. 34, pp. 129-153. Princeton Univ. Press, Princeton, N.J., 1956.
2. Blackwell, D., and Koopmans, L. On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **28**, 1011-1015 (1957).
3. Gilbert, E. J. On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **30**, 688-697 (1959).
4. Blackwell, D., Breiman, L., and Thomasian, A. J. Proof of Shannon's transmission theorem for finite-state indecomposable channels. *Ann. Math. Statist.* **29**, 1209-1220 (1958).
5. Shannon, C. E. The mathematical theory of communication. *Bell System Tech. J.* **27**, 379-423, 623-656 (1948).
6. Carlyle, J. W. Equivalent stochastic sequential machines. Tech. Rep. Ser. 60, Issue 415, Electronics Research Laboratory, Univ. of California, Berkeley, California, 1961.