IDENTIFICATION OF STATE-CALCULABLE FUNCTIONS OF FINITE MARKOV CHAINS[1]

By J. W. Carlyle

University of California, Los Angeles

1. Summary. If $\{f(X_n): n = 1, 2, \cdots\}$ is a finite-state function of a finite-state Markov chain $\{X_n\}$, it is known that the distribution of $\{f(X_n)\}$ is determined by the distribution of $(f(X_1), f(X_2), \cdots, f(X_K))$ for suitable K, and a finite construction utilizing only the latter distribution exists in special cases yielding the probability structure of a chain $\{X_n'\}$ and a function f' such that $\{f(X_n)\}$ and $\{f'(X_n')\}$ have the same distribution [7]. We obtain a finite construction when $(\{X_n\}, f)$ is such that if i is a state of $\{X_n\}$ and j is a state of $\{f(X_n)\}$, then there is at most one transition from i to the set $f^{-1}(j)$ and the distribution of X_1 assigns positive probability to at most one state in each set $f^{-1}(j)$. Such "state-calculability" is a rather severe, but natural, structural restriction subsuming certain cases not previously treated. The corresponding finite construction is very simple and directly related to the representation of any finite-state process $\{Y_n\}$ by a function of a (possibly countable-state) Markov chain.

2. Preliminaries. Let $\{Y_n : n = 1, 2, \cdots\}$ be a stochastic process with finite-state set J having D elements. Let J^* be the set of all y-sequences (finite sequences of elements of J). Letters s, t, u, v here denote y-sequences, while ϵ is a y-state (or sequence of length 1); $|s|$ is the length of s, and the sequence "s followed by t" is written st. Let $p(\cdot)$ be defined on J^* by

$$p(s) = P[(Y_1, Y_2, \cdots, Y_n) = s] \quad \text{when} \quad |s| = n. \tag{1}$$

We say that $\{Y_n\}$, or its probability function p, *is represented by a function of a Markov chain* if there exists a Markov chain $\{X_n\}$ with state set I and a function f from I to J such that $\{Y_n\}$ and $\{f(X_n)\}$ have the same distribution (it may be that $Y_n = f(X_n)$, but we are concerned here only with equality of probability laws). With $I = \{1, 2, \cdots\}$ finite or countable, we arrange the chain transition probabilities $P[X_2 = j \mid X_1 = i]$ and initial distribution $P[X_1 = i]$ in the usual way, in a matrix $M = (m_{ij})$ and row vector $m = (m_i)$ respectively, and we let either $(\{X_n\}, f)$ or (M, m, f) denote the representation of $\{Y_n\}$.

Suppose it is known that $\{Y_n\}$ admits such a representation, about which nothing is specified except that I can be taken to be finite with at most C elements (of course $C \geqq D$). Let $K = 2(C - D + 1)$; then the entire distribution of $\{Y_n\}$ is determined solely by the probabilities (1) for $n = K$, or equivalently the function p is determined by $[p]_K$, where $[p]_n$ denotes the restriction of p to argu-

Received 22 June 1966.

[1] This research was supported in part by the National Science Foundation under Grant GP-3008, and in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant 700-65.

ments of length $\leqq n$. This basic result (with restrictions on the chain $\{X_n\}$, including stationarity, which can be dropped) is due to Gilbert [7], who showed that K could not be reduced, and also obtained an explicit recurrence relation for $[p]_n$ based on $[p]_K$. However, explicit construction of a representation (M, m, f) from $[p]_K$ remains an open problem. (Blackwell and Koopmans [1] first considered "identifiability" problems of this nature, and obtained an upper bound for K.) Gilbert [7] found (M, m, f) when p satisfies a certain regularity condition and is representable as a function of a stationary irreducible aperiodic chain; the technique in [7] is still valid when the chain restrictions are removed, and also when the Markov matrix depends on a finite-state parameter [3], but the regularity condition is used in an essential way. We shall not need to review regularity here; it suffices to note that although structures (M, m, f) giving rise to regular $\{Y_n\}$ are "typical" in the sense of selection at random [7], it is easy to produce examples of structure classes containing both regular and nonregular members, particularly when restrictions are placed on the connection properties $\{X_n\}$ and their relationship to f. Our object here is to give a simple construction of (M, m, f) for one such class which arises rather naturally.

We note that other results (not directly relevant here) have appeared relating to identifiability [4] and to the determination of necessary and sufficient conditions that a probability function p be representable as a function of a finite-state Markov chain [5], [6], [8].

3. State-calculable representation. First we observe that any finite-state process $\{Y_n\}$ can be represented by a function of a Markov chain with at most countably many states; in particular we shall use the following representation. Let J^{*p} be the set of all $s \,\varepsilon\, J^*$ such that $p(s) > 0$. For $s \,\varepsilon\, J^{*p}$, define $p_s(\cdot)$ on J^* by $p_s(t) = p(st)/p(s)$. Let J^{*p} be partitioned into equivalence classes $E_1, E_2, \cdots$ by the equivalence relation

$$s \sim t \Leftrightarrow \text{final term of } s = \text{final term of } t,\ p_s(\cdot) = p_t(\cdot). \tag{2}$$

With a slight abuse of notation, let p_i be the probability function coinciding with p_s for all $s \,\varepsilon\, E_i$. Let ϵ_i be the final y-state common to each sequence in E_i; state i of our Markov chain is to be identified with class E_i, and we define f on the resulting state set I by

$$f(i) = \epsilon_i. \tag{3a}$$

We write $i \to j$ if there exists $s \,\varepsilon\, E_i$ such that $s\epsilon_j \,\varepsilon\, E_j$, and we say that i is *initial* if $\epsilon_i \,\varepsilon\, E_i$. It is easily verified that: $(\alpha) i \to j \Leftrightarrow s\epsilon_j \,\varepsilon\, E_j$ for all $s \,\varepsilon\, E_i$; (β) if $p_i(\epsilon) > 0$ then $i \to j$ for one and only one $j \,\varepsilon\, f^{-1}(\epsilon)$; (γ) if $p(\epsilon) > 0$ then i is initial for one and only one $i \,\varepsilon\, f^{-1}(\epsilon)$. Therefore, setting

$$\begin{aligned} m_{ij} &= p_i(\epsilon_j) \quad \text{when} \quad i \to j, \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{3b}$$

$$\begin{aligned} m_i &= p(\epsilon_i) \quad \text{if} \quad i \quad \text{is initial} \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{3c}$$

we obtain a Markov matrix M and a probability vector m. It follows immediately by inspection that (M, m, f) as defined by (3) is a representation of $\{Y_n\}$ as a function of a Markov chain.

The structure $(\{X_n\}, f)$ obtained from (3) is *state-calculable*; i.e., there exists a function $g: J \cup (I \times J) \to I$ such that, with probability one,

$$(4) \qquad g(f(X_1)) = X_1, \quad g(X_n, f(X_{n+1})) = X_{n+1}.$$

(The existence of g follows from (β) and (γ) preceding (3b).) Thus, using (3), any $\{Y_n\}$ can be represented as a state-calculable function of a Markov chain. Suppose that the required chain state set I is known to be finite; in the next section we show that the representation can then be constructed from $[p]_n$ for sufficiently large n.

First we establish two simple lemmas valid in the general case where I may be infinite. We say that $s, t \in J^{*p}$ are *n-equivalent* (otherwise *n-distinguishable*) if in (2) the requirement $p_s = p_t$ is replaced by $[p_s]_n = [p_t]_n$. Let Π_n be the partition of J^{*p} with respect to n-equivalence, and let Π denote the equivalence partition $\{E_1, E_2, \cdots\}$ already introduced.

LEMMA 1. *$\{\Pi_n : n = 1, 2, \cdots\}$ is a sequence of refinements, with common refinement Π. In fact, if n is such that $\Pi_n \neq \Pi$, then Π_{n+1} is a proper refinement of Π_n.*

PROOF. The first assertion is obvious. Now let n be such that $\Pi_n \neq \Pi$; then there exist s, t which are n-equivalent but not equivalent. Let s, t be $(n + k)$-equivalent and $(n + k + 1)$-distinguishable, and let uv be a minimal-length distinguishing sequence for s, t, with $|u| = k$, $|v| = n + 1$; then $p_s(u) = p_t(u) > 0$ and

$$(5) \qquad p_s(u)p_{su}(v) = p_s(uv) \neq p_t(uv) = p_t(u)p_{tu}(v),$$

showing that su and tu are $(n + 1)$-distinguishable. But equality prevails in (5) when v is replaced by any sequence of length n, so su and tu are n-equivalent. Hence, there is a class in Π_n which is refined in Π_{n+1}.

We say that E_i, or i, is *reachable in n steps* $(n \geqq 1)$ if there exists $s \in E_i$ with $|s| = n$. Let G_n be the set of all $i \in I$ reachable in at most n steps.

LEMMA 2. *$G_n \uparrow I$; in fact if n is such that $G_n \neq I$, then G_n is a proper subset of G_{n+1}.*

PROOF. Clearly $G_n \uparrow I$. To prove the remaining assertion, it suffices to show that when $G_n = G_{n+1}$ we have $G_{n+2} \subset G_{n+1}$. If $j \in G_{n+2}$ there exists s with $s\epsilon_j \in E_j$ and $|s| \leqq n + 1$. If $s < n + 1$ the conclusion is immediate. Let $|s| = n + 1$, and let $s \in E_i$; then $i \in G_{n+1} = G_n$, so there exists $t \in E_i$ with $|t| \leqq n$. Also $E_i \to E_j$ by construction. Thus, by (α) (above (3b)), $t\epsilon_j \in E_j$, or $j \in G_{n+1}$.

4. Finite construction. We say that $[Y_n]$, or p, is of *finite state-calculable type* if it is known that there is some representation as a state-calculable function of a finite-state Markov chain; if so, there are evidently many such representations (which we call *finite state-calculable representations*), and we now show that the "best," namely (3), can be obtained from $[p]_n$, for suitably large n, by a finite construction (i.e., a finite number of algebraic operations on the finite set of values of the function $[p]_n$).

THEOREM. *Let* $\{Y_n\}$ *be a process with D states and let p be defined by* (1). *Suppose it is known that* $\{Y_n\}$ *is of finite state-calculable type, and that C is an upper bound on the number of chain states necessary for a finite state-calculable representation. Then the unique minimal-state finite-calculable representation can be obtained through a finite construction which utilizes only* $[p]_K$ *for* $K = 2(C - D + 1)$.

PROOF. Let I, M, m, f refer to the representation (3) of Section 3 applied to the present $\{Y_n\}$. It is easy to see that I is finite with at most C states, that I has the least number of states among all finite state-calculable representations, and that such a minimal-state representation is unique except for permutations of the state set. It remains to show that those aspects of (3) which depend upon p in its entirety can be restated in terms of $[p]_K$ only, for the present $\{Y_n\}$. Using Lemma 1 we see that $\{\Pi_n\}$ here consists of a finite initial sequence of strict refinements with $\Pi_n = \Pi$ thereafter. Thus, since Π has at most C classes and Π_1 has at least D classes, $D + n - 1 \geqq C$ is sufficient to guarantee $\Pi_n = \Pi$, so that

$$(6) \qquad [p_s]_{C-D+1} = [p_t]_{C-D+1} \Rightarrow p_s = p_t .$$

Similarly, from Lemma 2, $\{G_n\}$ in the present case consists of a finite initial sequence which is strictly increasing with $G_n = I$ thereafter; G_1 has at least D elements and I has at most C elements, so $D + n - 1 \geqq C$ is sufficient for $G_n = I$. Therefore every $i \in I$ is reachable in at most $C - D + 1$ steps. This result, with (6), shows that one can determine the size of I, connections $i \rightarrow j$, and states ϵ_i (and then calculate m_{ij} and m_i) by inspection of the functions

$$(7) \qquad [p_s]_{C-D+1} : \text{ all } s \text{ of length} \leqq C - D + 1.$$

Evidently knowledge of $[p]_K$ is necessary and sufficient for evaluation of all of the functions in (7).

5. Remarks. The value of K in the above theorem cannot be reduced; simple examples where the strictly monotone portions of the sequences in Lemmas 1 and 2 are of maximal length can easily be exhibited by restricting attention to cases where M and m are degenerate (0 and 1 entries only). In this sense, the lemmas are extensions of well-known facts in deterministic automata theory, and similar results hold for random automata with a finite-state input [2]. The latter structures are called finite-state channels in information theory, and it is from this context [9] that the concept of state-calculability has been borrowed. Since functions of Markov chains serve as models for information sources, the result of Section 4 may be viewed as a method for deducing internal structure from externally "observable" characteristics for the class of finite state-calculable information sources.

Finally, we note that for any finite-state process $\{Y_n\}$ with probability function p and any positive integer k, the arguments of Sections 3 and 4 can be adapted in an obvious manner to yield a construction (based on $[p]_k$) of a finite state-calculable representation for a process $\{Z_n\}$ of finite state-calculable type such that $(Y_1, Y_2, \cdots, Y_k)$ and $(Z_1, Z_2, \cdots, Z_k)$ have the same distribution.

REFERENCES

[1] BLACKWELL, D. and KOOPMANS, L. (1957). On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **28** 1011–1015.

[2] CARLYLE, J. W. (1963). Reduced forms for stochastic sequential machines. *J. Math. Anal. Appl.* **7** 167–175.

[3] CARLYLE, J. W. (1964). On the external probability structure of finite-state channels. *Inform. Control.* **7** 385–397.

[4] DHARMADHIKARI, S. W. (1963). Functions of finite Markov chains. *Ann. Math. Statist.* **34** 1022–1032.

[5] DHARMADHIKARI, S. W. (1963). Sufficient conditions for a stationary process to be a function of a finite Markov chain. *Ann. Math. Statist.* **34** 1033–1041.

[6] DHARMADHIKARI, S. W. (1965). A characterization of a class of functions of finite Markov chains. *Ann. Math. Statist.* **36** 524–528.

[7] GILBERT, E. J. (1959). On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **30** 688–697.

[8] HELLER, A. (1965). On stochastic processes derived from Markov chains. *Ann. Math. Statist.* **36** 1286–1291.

[9] WOLFOWITZ, J. (1961). *Coding Theorems of Information Theory*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.