# REALIZATION OF STOCHASTIC SYSTEMS[1]

By Michael Arbib

*Stanford University*

## ABSTRACT

Heller has given necessary and sufficient conditions that a stochastic process be induced from a Markov chain. We consider a process induced by a Markov chain to be a probabilistic finite automaton with one input.

With each state of a probabilistic finite automaton, we may associate a function $p(u \mid v)$, which tells us the probability that, if we apply the input sequence $v$ to the machine started in the state, we should observe output sequence $u$. We give a necessary and sufficient condition that a function $p(u \mid v)$ be realizable as such an input-output function. Finally, we show Heller's result is extended by our condition.

Heller (1965) has given a necessary and sufficient condition that a stochastic process be induced from a Markov chain.

Let $S$ be a finite set and let $\mathcal{P}(S)$ be the set of maps $S^* \to R$ (the dual of $S^*$) (where, for any set $S$, we write $S^*$ for the set of finite sequences of elements from $S$). Then $p \varepsilon \mathcal{P}(S)$ is a *stochastic process* if it satisfies

(a) $p(S^*) \subseteq [0, 1]$,

(b) $\sum_{s \varepsilon S} p(s) = 1$,

(c) $\sum_{s \varepsilon S} p(s_1, \cdots, s_n, s) = p(s_1, \cdots, s_n)$.

On recalling the notion of probabilistic finite automaton, we shall see that it is an appropriate generalisation of the notion of "stochastic process induced by a Markov chain."

A *probabilistic finite automaton* (pfa) is a quadruple $(X, Y, Q, P)$ where

- $X$ is a finite set: the set of inputs,
- $Y$ is a finite set: the set of outputs,
- $Q$ is a finite set: the set of states, and
- $P$ is a conditional probability function $P(q', y \mid q, x)$.

We interpret $P(q', y \mid q, x)$ as the probability that if the machine is started in state $q \varepsilon Q$ and input $x \varepsilon X$ is applied, then the output should be $y \varepsilon Y$ and the next state should be $q' \varepsilon Q$.

Rather than concentrating on the finite set $Q$, it is more natural to consider $\Pi$, the set of probability distributions on $Q$, augmented by a separate state called 0, as state space. We may then define a next state function by

---

Received 31 March 1966; revised 4 November 1966.

[1] This research was supported in part by the U. S. Air Force under Grant No. AF-AFOSR-1198-67.

$$\tau_P[\pi, (y \mid x)]_j = \sum_i \pi_i \cdot P(q_j, y \mid q_i, x) / \sum_j P(q_j, y \mid q_i, x) \quad \text{if } \sum_j P(q_j, y \mid q_i, x) \text{ is } > 0$$
$$= 0 \qquad \text{otherwise.}$$

The *input-output function* of a state $\pi$ is defined to be the function

$$p_\pi(y_1, \cdots, y_n \mid x_1, \cdots, x_n) = \sum_{q_k \varepsilon Q, 1 \leq k \leq n} \pi_{q_i} \prod_{k=1}^{n} P(q_{k+1}, y_k \mid q_k, x_k),$$

and is the probability, given initial distribution state $\pi$ and input sequence $x_1, \cdots, x_n$, that the output sequence be $y_1, \cdots, y_n$.

We say states $\pi$ and $\pi'$ are *equivalent* if they have the same input-output functions, i.e., $p_\pi(\cdot|\cdot) \equiv p_{\pi'}(\cdot|\cdot)$. A pfa is in reduced form if no distinct states are equivalent.

DEFINITION. A *state-output* pfa is one in which the output $y$ is a function $f$ of the state $q'$, i.e., $P(q', y \mid q, x) = 0$ unless $y = f(q')$.

THEOREM 1. *Every* pfa *is equivalent to a state-output* pfa.

PROOF. Just take new finite state-set $Q \times Y$, and use the transition function

$$\tilde{P}((q, y'), y'' \mid (q, y), x) = 0 \qquad \text{if } y' \neq y''$$
$$= P(q, y' \mid q, x) \qquad \text{if } y' = y'' \qquad \text{QED}$$

This is, of course, a standard construction.

A stochastic process which is, in our terminology, an input-output function of a state-output, one input pfa, is clearly what is usually called "a stochastic process induced by a Markov chain."

We must now generalise the notion of stochastic process to yield our notion of "stochastic system" in such a way that a stochastic system with one input reduces to a stochastic process.

Let $X$ and $Y$ be finite sets, $R$ the real line, and $\mathcal{P}(X, Y)$ the space of all functions p: $(X \times Y)^* \rightarrow R$. For $p \varepsilon \mathcal{P}(X, Y)$ we write

$$p(y_1, \cdots, y_n \mid x_1, \cdots, x_n) \quad \text{for} \quad p((x_1, y_1), \cdots, (x_n, y_n)).$$

DEFINITION. A function $p \varepsilon \mathcal{P}(X, Y)$ is called a *stochastic system* if it satisfies the three conditions

(S1) $p((X \times Y)^*) \subseteq [0, 1]$,

(S2) $\sum_{y \varepsilon Y} p(y \mid x) = 1$,

(S3) $\sum_{y \varepsilon Y} p(y_1, \cdots, y_n, y \mid x_1, \cdots, x_n, x) = p(y_1, \cdots, y_n \mid x_1, \cdots, x_n)$

for all $x \varepsilon X$.

If we let $X$ have but one element, and then identify $X \times Y$ with $Y$, which we relabel $S$, conditions (S1–S3) do indeed reduce to the conditions (a)–(c) for a stochastic process. Thus, we may inded think of a stochastic process as a stochastic system in which we have no control over the input.

Note that a stochastic system is *not* a stochastic process in which $X \times Y$ replaces $S$. We denote by $\mathcal{P}_s(X, Y)$ the convex subset of $\mathcal{P}(X, Y)$ comprising the stochastic systems.

Clearly any input-output function for a state of a pfa is a stochastic system. Our central aim in this paper is to characterize such stochastic systems, thus generalizing Heller's result.

Our study is prompted by the work of Carlyle, additions to which have been made by Bacon, Even, and Page. These authors have only studied stochastic systems in which the assumption is made *ab initio* that they come from a pfa. The removal of this assumption parallels the work of Raney for the deterministic case.

Henceforth we shall restrict our attention to $\mathcal{P}_s(X, Y)$, but the reader should note that many of our notions and proofs carry over almost unchanged to the general case of $\mathcal{P}(X, Y)$.

Let $\mathcal{P}^1(X, Y)$ be the set of stochastic systems restricted to $X \times Y$, which is

$$\{p \mid X \times Y : P \varepsilon \mathcal{P}_s(X, Y)\}.$$

A *restricted stochastic system* (RSS) is a quintuple $\mathcal{R} = (X, Y, \Pi, \tau, \delta)$ where, setting $Z = X \times Y$, we have

$\Pi$ is a convex set to which a 0 is adjoined

$\tau : \Pi \times Z \rightarrow Z$ is such that if $\pi_1$ and $\pi_2$ are in $\Pi$, then for all convex combinations $a_1\pi_1 + a_2\pi_2$, we have $\tau(a_1\pi_1 + a_2\pi_2\, z)$ is a convex combination of the non-zero images among $\tau(\pi_1, z)$ and $\tau(\pi_2, z)$—and 0 is there are none

$\delta : \Pi - \mathcal{P}^1(X, Y) \cup \{0\}$. with $\delta(0) = 0$.

Given an RSS, $\mathcal{R}$, we may associate with each "state" $\pi \varepsilon \Pi$ an "input-output" function $\hat{p}_\pi$ by the equation:

$$\hat{p}_\pi(y_1, \cdots, y_n \mid x_1, \cdots, x_n) = \prod_{k=0}^{n-1} \delta(\pi_k)[y_{k+1} \mid x_{k+1}]$$

where $\pi_0 = \pi$, $\pi_{k+1} = \tau[\pi_k, (y_{k+1} \mid x_{k+1})]$, $0 \leqq k < n$.

Definition. An RSS $\mathcal{R}$ is a *realization* of the function $p \varepsilon \mathcal{P}_s(X, Y)$ if there is a state $\pi$ of $\mathcal{R}$ whose input-output function $\hat{p}_\pi$ is equal to $p$.

An RSS is said to be in *reduced form* if $\hat{p}_{\pi_1} = \hat{p}_{\pi_2} \Rightarrow \pi_1 = \pi_2$, and all states are *reachable* from some state i.e., $\exists \pi \varepsilon \Pi$ such that $\Pi = \tau(\pi, Z^*)$.

The reader should note that any state $\pi$ of a pfa $M$ induces a realization $(X, Y, \Pi, \tau, \delta)$ with $\delta(\pi')(y \mid x) = p_{\pi'}(y \mid x)$.

Given any pfa $M$, we may map its distribution states $\pi$ into elements $p_\pi$ of $\mathcal{P}_s(X, Y)$. It is clear that if $M$ is in minimal-state form (Bacon, 1959),then no image of any pure state can be expressed as a linear combination of the images of other pure states.

What is surprising is that distinct *distribution* states of a minimal-state machine may map to the same stochastic system. A. Paz (personal communication) gave me the following example. Consider the 4-state deterministic machine for which all inputs cause a transition to the 4th state, but the respective outputs for inputs 0 and 1 are 1 and 1 for state $q_1$, 0 and 0 for state $q_2$, 1 and 0 for state $q_3$, and 0 and 1 for state $q_4$. If we consider this, as a stochastic machine, it is clearly in minimal-state form. However, $\pi_1 = (\frac{1}{2}, \frac{1}{2}, 0, 0)$ and $\pi_2 = (0, 0, \frac{1}{2}, \frac{1}{2})$ have $p_{\pi_1} = p_{\pi_2}$.

LEMMA. *Let* $p \in \mathcal{P}_s(X, Y)$ *be induced by a state* $\pi$ *of a* pfa $M$. *Let* $M'$ *be the sub-automaton of* $M$ *comprising the states reachable from* $\pi$ *with positive probability. Then any reduced realization of* $p$ *may be obtained from* $M'$, *by mapping its distribution states into the corresponding elements of* $\mathcal{P}_s(X, Y)$.

THEOREM 2. *Every function in* $\mathcal{P}_s(X, Y)$ *has a realization.*

PROOF. We define an action of $Z^*$ on $\mathcal{P}_s(X, Y)$ as follows:

For $z = (y \mid x) \in Z^*$ and $p \in \mathcal{P}_s(X, Y)$, we define the function $p \cdot z$ by the equation

$$[p \cdot z](y_1 \mid x_1) = p(yy_1 \mid xx_1)/p(y \mid x) \quad \text{if } p(y \mid x) > 0$$
$$= 0 \quad \text{if not.}$$

The reader may readily verify that

(i) $p \cdot z$ is in $\mathcal{P}_s(X, Y) \cup \{0\}$,

(ii) for all $\tilde{z}$ and $\bar{z}$ in $Z^*$, $[p \cdot \tilde{z}] \cdot \bar{z} = p \cdot (\tilde{z}\bar{z})$,

(iii) $z$ carries convex combinations of points into (usually different) convex combinations of their *non-zero* images—and into 0 if there are none.

$$\textstyle\sum_j a_j p_j \cdot z = \sum_j (a_j p_j(z)/\sum_k a_k p_k(z)) p_j \cdot z$$

unless all $p_j \cdot z$ are 0.
It then follows that $(X, Y, \mathcal{P}_s(X, Y), \hat{\tau}, \hat{\delta})$ is a realization of $p \in \mathcal{P}_s(X, Y)$ as soon as we take

$$\hat{\tau}(p', z) = p' \cdot z,$$
$$\hat{\delta}(p') = p' \mid Z. \qquad \text{QED}$$

THEOREM 3. *Every* $p \in \mathcal{P}_s(X, Y)$ *has a reduced realization. Furthermore this realization is unique up to isomorphism,* i.e., *if* $(X, Y, \Pi_1, \tau_1, \delta_1)$ *and* $(X, Y, \Pi_2, \tau_2, \delta_2)$ *are both reduced realizations of* $p$, *then there is a one-to-one correspondence* $\theta : \Pi_1 \to \Pi_2$ *such that*

$$\tau_2(\theta\pi, z) = \theta\tau_1(\pi, z),$$
$$\delta_1(\pi) = \delta_2(\theta\pi).$$

PROOF. To get a reduced realization, we merely replace $\mathcal{P}_s(X, Y)$ in the realization of Theorem 1 by $p \cdot Z^*$. Uniqueness follows from the observation that any realization $(X, Y, \Pi, \tau, \delta)$ enjoys the property that for all $z \in Z^*$,

$$\delta(\tau(\pi, z)) = (p \cdot z) \mid Z$$

(extending $\tau$ to a function $\Pi \times Z^* \to \Pi$ in the natural fashion). QED

THEOREM 4.[2] *The stochastic system* $p$ *is induced by a state of a* pfa *iff the function* $p$ *is contained in a polyhedral convex set whose union with* 0 *is closed under the*

[2] An earlier draft of this paper only considered closure under an action of $X^*$. I am grateful to Alex Heller and the referee for drawing the insufficiency of this condition to my attention.

*action of $Z^*$ in $\mathcal{P}_s(X, Y)$*, i.e., iff $\exists$ *a finite set* $\{q_1, \cdots, q_n\} \subseteq \mathcal{P}_s(X, Y)$ *of non-zero $q_i$'s such that p and each non-zero $q_i \cdot z$ may be expressed as a convex combination* $\sum \lambda_i q_i$ *for suitable constants* $\lambda_i (0 \leqq \lambda_i \leqq 1, \sum \lambda_i = 1)$.

PROOF OF THEOREM 4. The "only if" part is obvious. For the "if" part, suppose that $q_1, \cdots, q_n$ have the stated properties. We exhibit a pfa $M$ with finite state set $q_1, \cdots, q_n$ and for which $\pi = (\pi_1, \cdots, \pi_n)$ implies that the corresponding $p_\pi$ for $M$ is just $\sum \pi \hat{p}_{q_i}$. Our task is to find matrices

$$P(q_j, y \mid q_i, x) = m_{ij}(y \mid x)$$

which describe the transition and output probabilities. To do this we must simply solve the equations

$$\delta(q_i)[y \mid x] = \textstyle\sum_j m_{ij}(y \mid x), \tag{1}$$

$$q_i \cdot (y \mid x) = \textstyle\sum_j m_{ij}(y \mid x) q_j / \sum_j m_{ij}(y \mid x) \quad \text{if} \quad \sum_j m_{ij}(y \mid x) > 0 \tag{2}$$

$$= 0 \qquad \text{otherwise.}$$

Multiplying the coefficient of $q_i$ in (2) by the $\sum_j m_{ij}(y \mid x)$ of (1), we determine $m_{ij}(y/x)$ for the given set $\{q_1, \cdots, q_n\}$. The example of Paz, cited above, indicates that the $m_{ij}$ need not be unique. QED

Our only task now is to recall enough of Heller's terminology to show that his condition is indeed a restriction of ours. Recall the conditions (a)–(c) on a stochastic process $p$.

We may extend $p$ to $A_S$, the free associative $R$-algebra generated by $S$ by $p(\sum_{x \varepsilon S^*} f(x)x) = \sum_{x \varepsilon S^*} f(x)p(x)$, and setting $p(1) = 1$.

Note that $A_S$ is the *dual* of $S^*$, but we keep $\mathcal{P}(S)$ and $A_S$ distinct.

Denote by $P_S$ the coordinate cone of $A_S$ consisting of polynomials with non-negative coefficients, and let $\sigma = \sum_{x \varepsilon S} x$. Then an $R$-linear $p: A_S \to R$ is a stochastic process if the following conditions hold

(P0) $p(1) = 1$,

(P1) $p(P_S) \subseteq [0, \infty]$,

(P2) for all $\xi \varepsilon A_S$, $p(\xi\sigma) = p\xi$.

Then a realization of a stochastic process $p$ is a quintuple $(S, \Pi, \tau, \delta)$ is a RSS which has a state $\pi$ with $p_\pi = p$.

Heller introduces the notion of a stochastic $S$-module ($sS$-module) as follows:

Let $L$ be a right $A_S$-module[3], $l_0 \varepsilon L$, and $q: L \to R$ (linear). Then $p\xi = q(l_0\xi)$ defines a linear $p: A_S \to R$. Call $(L, q, l_0)$ a *sS-module* if

(i) $ql_0 = 1$,

(ii) $q(l_0 P_S) \subseteq [0, \infty]$,

(iii) for all $\xi \varepsilon A_S$, $q(l_0\xi(\sigma - 1)) = 0$. Then clearly $p$ is a stochastic pro-

[3] Heller uses *left* modules, but I find right modules easier to use.

cess if $(L, q, l_0)$ is an $sS$-module—we then say that $p$ is the stochastic process *associated* with $(L, q, l_0)$.

The reduced realization of a stochastic process $p$ is obtained from the reduced $sS$-module with which $p$ is associated. We make $\mathcal{P}(S)$ into an $A_S$-module by setting

$$ls(y) = l(sy) \quad \text{for} \quad l \in \mathcal{P}(S),\ s \in S^*$$

and we extend this by linearity

$$lf = \textstyle\sum_{s \in S^*} f(s) ls \quad \text{for} \quad f \in A_S .$$

The reduced module for $p$ is just $(pA_S, q, p)$ where $q(l) = l(1)$. The reduced realization for $p$ is just $(S, pS^*, \tau, \delta)$ where $l(\pi, \xi) = \pi\xi$, $\delta(\pi) = \pi \mid \delta$, i.e., Heller has just taken all linear combinations of the functions rather than just convex combinations.

Lemma. *Every stochastic sS-module defines a realization of the associated stochastic process p as follows:*

*Take* $\Pi = L$.

$$\tau(l, \xi) = l\xi,$$
$$\delta(l)(s) = q(ls),$$
$$\pi = l_0 .$$

Heller calls an $sS$-module $(L, q, l_0)$ *reduced* if (i) $L$ is cyclic with generator $l_0$, i.e., $l_0 A_s = L$ and (ii) $L$ has no nonzero submodule $L'$ with $q(L') = 0$.

By a *cone* in a real vector space $V$ we mean a union of rays from the origin. A convex cone $C$ is *strongly convex* if it contains no line through the origin; and is *polyhderal* if it is the convex hull of the union of finitely many rays.

We may now state Heller's main theorem (his Theorem 5.1):

Theorem 5. *Let $(L, q, l_0)$ be a reduced sS-module. The associated stochastic process is induced from a Markov chain if there is a cone $C \subset L$ such that*

(i) $l_0 \in C$,

(ii) $q(C) \subset [0, \infty]$,

(iii) *C is invariant under $P_S$*, i.e., $CP_S \subseteq C$,

(iv) *C is strongly convex and polyhedral.*

Taking note of Theorem 1, and our discussion of the relation between the reduced $sS$-module and the reduced realization, we see that this is an immediate consequence of Theorem 4.

## REFERENCES

[1] Bacon, Glenn C. (1964). Minimal-state stochastic finite state systems. *IEEE Trans.* **CT-11** 307–308.

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

[3] Carlyle, J. W. (1965). State-calculable stochastic sequential machines, equivalences,

and events. *Proc. IEEE Sixth Annual Symposium on Switching Circuit Theory and Logical Design*.

[4] Even, Shimon (1965). Comment on the minimization of stochastic machines. *IEEE Trans.* **EC-14** 634–637.

[5] Heller, Alex (1965). On stochastic processes derived from Markov chains. *Ann. Math. Statist.* **36** 1286–1291.

[6] Page, Carl V. (1965). Equivalences between probabilistic machines. Technical Report 03105-41-T, University of Michigan.

[7] Raney, G. N. (1958). Sequential functions. *J. Assoc. Comp. Machinery* **5** 177–180.