KYBERNETIKA ČÍSLO 6, ROČNÍK 2/1966

Theory of Stochastic Automata*

PETER H. STARKE

This paper gives, summarizing the papers [6]–[9], a short introduction to the theory of stochastic automata which has been developed within the last two years.

The theory of abstract deterministic automata was developed in a relatively short time from practical points of view to a mathematical theory which has found a great number of applications. Such applications are for instance:

applications to the construction and analysis of discrete-working information-processing systems,
applications to Linguistics,
applications to the foundations of Mathematics.

But there is a great number of essential cybernetical problems which can not be treated within the framework of this theory. For instance such questions which concern the notion of learning, especially the imitation of the formation of a conditioned reflex by an animal. Problems of the accomodation of a system to its environment are treated within the theory of game-playing machines too.

If we want to imitate by a machine the formation of a conditioned reflex, we need a system which reacts probabilistically at least in a given interval of time, namely the time of that formation. Whithin that interval of time there is only some probability that the animal reacts, when the conditioned stimulus is presented but no sureness.

Therefore it is a presupposition for the development of a theory of learning automata that a theory of probabilistic (or as we call them – stochastic) automata should be worked out. Because the latter is missing, in treating problems of the mentioned type the authors are urged to apply ad-hoc-constructions of stochastic automata ([1], [2], [3]).

In my lecture I would like to report on some of the basic definitions and results of a theory of stochastic automata, which has been developed within the last two years ([6], [7], [8]).

It seems to be clear, how to define the notion *stochastic automaton*. The definition should be established in such a manner, that the usual deterministic automata in some sense are found out as special cases of stochastic automata.

* Presented at the *Second Conference on Cybernetics*, Prague, November 16th–19th, 1965.

Deterministic automata are described usually by a quintupel $[X, Y, Z, f, g]$ of three non-empty sets, namely the set X of all input signals x, the set Y of all output signals y and the set Z of all states z of the system, and two functions, namely the transition function f, which maps the set $Z \times X$ (of all ordered pairs $[z, x]$ with z from Z and x from X) into the set Z, and the output function g, which maps the same set into the set Y. The system $[X, Y, Z, f, g]$ works within a discrete time scale $t = 1, 2, 3, \ldots$, so that the so-called canonical equations hold:

$$z_{t+1} = f(z_t, x_t)\,,$$
$$y_t = g(z_t, x_t)$$

(z_t is the state of the system at time t, x_t is the input signal at time t and so on).

We call each pair $[z, x]$ from $Z \times X$ a situation of the automaton $[X, Y, Z, f, g]$. We see from the definition, that at each time t the behavior of the deterministic automaton is determined by the situation at time t.

In a stochastic automaton the state z_{t+1} at time $t + 1$ is in general not uniquely determined by the situation at time t, but for each state z from Z there is a probability fixed by the situation at time t, that $z = z_{t+1}$; that means that z is the state at time $t + 1$.

So it seems to be natural to substitute for the transition function f a function F, which maps the set $Z \times X$ of all situations into the set of all distribution-laws on the set Z. To the situation $[z_t, x_t]$ at time t corresponds the distribution-law $F[z_t, x_t]$ and the value $F[z_t, x_t](z^*)$ of that function is the probability that z^* is the state of the system at time $t + 1$.

Analogously, one can substitute a function G for the output function g (cf. [5]).

But one of our results is that the type $[X, Y, Z, F, G]$ of stochastic automata is not the most general one.

To get this general notion of a stochastic automaton, we start with an other definition of deterministic automata which is equivalent to that one mentioned above.

$[X, Y, Z, h]$ is a *deterministic automaton* if and only if X, Y, Z are non-empty sets and h is a function, which maps the set $Z \times X$ into the set $Y \times Z$. The system works within a discrete time scale, so that for each $t = 1, 2, 3, \ldots$ the equation

$$[y_t, z_{t+1}] = h(z_t, x_t)$$

holds.

Corresponding to this definition we define:

Definition 1. $[X, Y, Z; H]$ is a *stochastic automaton* if and only if:

1. X, Y, Z are non-empty at most countable sets (X is the set of all input signals, Y the set of all output signals and Z the set of all states of the system).
2. H is a function which is defined on the set $Z \times X$ of all situations and which has

for values distribution-laws on $Y \times Z$ (so that for each $[z, x]$ from $Z \times X$ holds: $\sum_{y \in Y, z^* \in Z} H[z, x](y, z^*) = 1$).

3. The system $[X, Y, Z; H]$ works within a discrete time scale $t = 1, 2, 3, \ldots$, so that for each t holds: If $[z_t, x_t]$ is the situation at time t, then the real number $H[z_t, x_t](y, z^*)$ is the probability, that z^* is the state of the system at time $t + 1$ and y is the output signal of the system at time t.

The limitation of our investigations to automata, which have at most countable sets of input signals, output signals and states, is motivated by our desire to avoid an application of measure- and integration-theory.* Besides, the main point of the theory is naturally the investigation of finite stochastic automata, that means automata which have only a finite number of input signals, output signals and states.

It is clear from the definition that stochastic automata don't change their behavior with the course of time. If the situations at time t_1 and at time t_2 are the same, the probabilities of the next state and the outputsignal are the same — that means, H doesn't depend on t explicitly.

By $W(X)$ we denote the set of all finite sequences $p = x_1 x_2 \ldots x_n$ of elements of X — the empty sequence is denoted by e. If is obviously, that the set $W(X)$ with the operation of juxtaposition is a free semi-group.

The length of a sequence p (the finite ordinal number of its domain) we denote by $l(p)$.

The first question, to which definition 1 leads immediately, is the question for the *behavior* of a stochastic automaton $\mathfrak{A} = [X, Y, Z; H]$, which is in a given initial state z_1 (at time 1) and into which is fed a sequence $p = x_1 x_2 \ldots x_n$ of input signals from X. That means, given a sequence $q = y_1 y_2 \ldots y_m$ from $W(Y)$, what is the probability that the automaton yields the sequence q under the mentioned conditions?

We define a function V_H which for each z_1 from Z and p from $W(X)$ yields a distribution-law on $W(Y)$, so that $V_H[z_1, p](q)$ is the probability in question:

$$V_H[z_1, p](q) = \begin{cases} 1, \text{ if } p = e \text{ and } q = e \\ \sum_{z_2 \in Z} H[z_1, x_1](y_1, z_2) \sum_{z_3 \in Z} H[z_2, x_2](y_2, z_3) \ldots \\ \ldots \sum_{z_{n+1} \in Z} H[z_n, x_n](y_n, z_{n+1}), \text{ if } p = x_1 x_2 \ldots x_n \text{ and } q = y_1 y_2 \ldots y_n \\ \qquad \text{(that means, if } l(p) = l(q) = n > 0) \\ 0, \text{ in all other cases.} \end{cases}$$

It can be seen easily that V_H has the required property.

* Meanwhile that limitation has been eliminated; cf: H. Thiele, P. H. Starke: Zufällige Zustände in stochastischen Automaten. Forthcoming in: Elektronische Informationsverarbeitung und Kybernetik.

We denote by xp the sequence which is constructed by juxtaposition from a signal x and a sequence p.

One can easily prove that for each $z \in Z$, $p \in W(X)$, $q \in W(Y)$, $y \in Y$, $x \in X$ the following equations hold:

$$V_H[z, xp](yp) = \sum_{z^* \in Z} H[z, x](y, z^*) \cdot V_H[z^*, p](q), \quad \sum_{q \in W(Y)} V_H[z, p](q) = 1 .$$

Some more definitions:

Let $\mathfrak{A} = [X, Y, Z; H]$ be an arbitrary stochastic automaton.

The function

$$F[z, x](z^*) \underset{\mathrm{df}}{=} \sum_{y \in Y} H[z, x](y, z^*)$$

is called the *transition-function* of $\mathfrak{A}$.

$F[z, x](z^*)$ is the probability that the next state is z^*, if the situation is $[z, x]$. It is sometimes essential to know the probability that the state of the automaton at time $t + n$ is z^*, if the automaton is in state z at time t and the sequence $p = x_1 x_2 \dots x_n$ is fed into the automaton during the time-interval $t, t + 1, \dots \dots, t + n - 1$. Therefore we enlarge the domain of F onto $Z \times W(X)$, so that $F[z, p](z^*)$ is the considered probability.

Then for each $p, r \in W(X)$; $z, z^* \in Z$

$$F[z, pr](z^*) = \sum_{z' \in Z} F[z, p](z') \, . \, F[z', r](z^*)$$

holds.

Further we define:

The function

$$G[z, x](y) \underset{\mathrm{df}}{=} \sum_{z^* \in Z} H[z, x](y, z^*)$$

is called the *output-function* of the stochastic automaton $\mathfrak{A}$.

We call a stochastic automaton $\mathfrak{A}$ *Z-determinated* if and only if for each situation $[z, x]$ there exists exactly one state z^*, so that

$$F[z, x](z^*) > 0 \quad (\text{that means:} \quad F[z, x](z^*) = 1),$$

and we call it *Y-determinated* if and only if for each situation $[z, x]$ there exists exactly one y, so that $G[z, x](y)$ is greater than 0. $\mathfrak{A}$ is called *determinated* if and only if it is Z-determinated and Y-determinated.

It is clear that, if $\mathfrak{A}$ is a determinated stochastic automaton, we can build up a deterministic automaton, which – with probability 1 – does the same.

$\mathfrak{A}$ is called a stochastic *Mealy-automaton* if and only if

$$H[z, x](y, z^*) = G[z, x](y) \cdot F[z, x](z^*)$$

always holds.

That means, that the chance events "the next state is z^*" and "the output signal is y" are always independent.

$\mathfrak{A}$ is called a stochastic *Moore-automaton* if and only if there exists a function M which is defined on Z and which has for values distribution-laws on Y, so that

$$H[z, x]\,(y, z^*) = F[z, x]\,(z^*)\,.\,M[z^*]\,(y)$$

always holds.

One can easily prove that each stochastic automaton which is Z-determinated or Y-determinated is a stochastic *Mealy-automaton.*

After these preparations now we turn to the problem of equivalence. Usually two systems are called equivalent, if we can substitute one for the other within all connections, in which only the external behavior of these systems is concerned.

Therefore we define the equivalence only for such automata, for which the sets of input signals on the one hand and the sets of output signals on the other hand are identical.

Definition 2. Let $\mathfrak{A} = [X, Y, Z; H]$ and $\mathfrak{A}' = [X, Y, Z'; H']$ be arbitrary stochastic automata, $z \in Z$, $z' \in Z'$ arbitrary states.

We say, that the state z is *equivalent* to the state z' $(z \sim z')$, if and only if for all $p \in W(X)$, $q \in W(Y)$

$$V_H[z, p]\,(q) = V_{H'}[z', p]\,(q)$$

holds.

We call the automata $\mathfrak{A}$, $\mathfrak{A}'$ *equivalent* if and only if for each state z from Z there exists an equivalent state z' from Z' and for each state z' from Z' there exists an equivalent state z from Z.

One can show, that the following holds:

For each stochastic automaton $\mathfrak{A} = [X, Y, Z; H]$ *there exists an equivalent stochastic Moore-automaton* $\mathfrak{B} = [X, Y, Z^*; F^*, M^*]$ *where the function* M^* *is determinated*, that means, that for each $z^* \in Z^*$ there exists (exactly) one y, so that $M^*[z^*]\,(y) = 1$.

In the process of *reduction* of a stochastic automaton $\mathfrak{A} = [X, Y, Z; H]$, that means in the construction of a stochastic automaton $\overline{\mathfrak{A}}$ which is equivalent to $\mathfrak{A}$ and in which each two different states are non-equivalent, some difficulties arise.

If $\bar{Z}$ denotes the set of all classes, into which the set Z is partioned by the relation of equivalence of states, and if $\mathfrak{A}^* = [X, Y, Z^*; H^*]$ is a reduced stochastic automaton, which is equivalent to $\mathfrak{A}$, then there exists obviously an one-to-one mapping of $\bar{Z}$ onto Z^*. Therefore we can take the set $\bar{Z}$ as set of states of a reduced automaton of $\mathfrak{A}$.

In order to build up such a reduced automaton $[X, Y, \bar{Z}; \bar{H}]$, we have to define $\bar{H}$. But this is possible in exactly one manner if and only if the following condition (**R**) holds:

(**R**) *For each* $z, z' \in Z$, $x \in X$, $y \in Y$, $[z_1] \in \bar{Z}$ ($[z_1]$ is that class which contains z_1) *it holds*:

If $z \sim z'$, *then*

$$\sum_{z'' \in [z_1]} H[z, x](y, z'') = \sum_{z'' \in [z_1]} H[z', x](y, z'') .$$

If (R) doesn't hold and $\mathfrak{A}$ is not reduced, then there exists an uncountable set of reduced and non-isomorphic stochastic automata, which are equivalent to $\mathfrak{A}$ (cf. [9]).

The condition (R) holds for each Z-determinated stochastic automaton. In general it is open, whether (R) holds or not.

Now we turn to a consideration of the *mappings* realized by stochastic automata, if we fix the initial state. Such a mapping Φ associates with each pair $[p, q]$ from $W(X) \times W(Y)$ a real number λ between 0 and 1 inclusively. It is clear, $\lambda = \Phi(p, q)$ is the probability that the automaton yields the sequence q, if the sequence p is fed into it.

Definition 3. Φ is a *stochastic operator* on $[X, Y]$ if and only if

1. X and Y are non-empty at most countable sets,
2. for each p from $W(X)$

$$\sum_{q \in W(Y)} \Phi(p, q) = 1 .$$

holds.

Let be Φ an arbitrary stochastic operator on $[X, Y]$. We call Φ *generated by the state* $z \in Z$ within the automaton $[X', Y', Z; H]$ if and only if the following conditions hold:

1. $X \subseteq X'$, $Y \subseteq Y'$.
2. for each $p \in W(X)$, $q \in W(Y)$: $V_H[z, p](q) = \Phi(p, q)$.

Let be Φ an arbitrary stochastic operator on $[X, Y]$ and $\mathfrak{A} = [X', Y', Z; H]$ an automaton, within which Φ can be generated. One can prove, that under these conditions there exists a stochastic automaton $[X, Y, Z^*; H^*]$ (which has the set X as set of input signals and the set Y as set of output signals), by which Φ can be generated.

If Φ is a stochastic operator on $[X, Y]$ which can be generated within a stochastic automaton, the following two conditions are satisfied:

$$\text{(I)} \qquad \text{If } l(p) \neq l(p), \quad \text{then} \quad \Phi(p, q) = 0 ;$$

$$\text{(II)} \qquad \sum_{y \in Y} \Phi(px, qy) = \Phi(p, q) .$$

Stochastic operators, which fulfil these two conditions, we call *sequential*.

Theorem. *A stochastic operator is sequential if and only if it can be generated within a stochastic automaton.*

In order to prove this theorem it is convenient to introduce the notion of a *state of a sequential stochastic operator*.

For each pair $[p, q]$ from $W(X) \times W(Y)$, such that $\Phi(p, q) > 0$, we define the stochastic operator

$$\Phi_{p,q}(r, s) = \frac{\Phi(pr, qs)}{\Phi(p, q)}$$

(r from $W(X)$, s from $W(Y)$ arbitrarily).

Each state of a sequential stochastic operator is a sequential stochastic operator. Each sequential stochastic operator can be generated by a stochastic automaton, which has as many states as the considered operator.

But the number of states of a sequential stochastic operator is not bounded by the minimal number of states of stochastic automata, by which this operator can be generated. There are stochastic operators with an infinite number of states, which can be generated by a finite stochastic automaton (cf. [8]).

Further there are sequential stochastic operators, which can not be generated by any stochastic Mealy-automaton. Thus the capabilities of stochastic Moore-automata are greater than those of stochastic Mealy-automata.

Finally let us consider stochastic events.

Definition 4. φ is a *stochastic event* on X if and only if

1. X is a non-empty at most countable set
2. φ is a function, which maps the set $W(X) \setminus \{e\}$ into the real interval $\langle 0, 1 \rangle$.

We may consider the value $\varphi(p)$ e.g. as the probability, that the sequence p occurs in some observation.

φ is called *represented by* $[z, S]$ within the stochastic automaton $[X, Y, Z; H]$ ($z \in Z, S \subseteq Z$) if and only if for all non-empty sequences p from $W(X)$

$$\varphi(p) = \sum_{z^* \in S} F[z, p](z^*)$$

holds.

That means, $\varphi(p)$ is the probability of finding the automaton in a state of S, if it is started in state z and if the sequence p was fed into the automaton.

One can show that each stochastic event is representable. The problem of characterizing the class of all stochastic events which can be represented by finite stochastic automata is solved only in a special case, when the range of the event φ is finite (that means, that φ has only a finite number of values).

Let X be a finite non-empty set, φ a stochastic event on X and λ a real number between 0 and 1 inclusively. Then we define the set

$$E_{\varphi=\lambda} \underset{\mathrm{df}}{=} \{p \mid p \in W(X) \setminus \{e\}, \quad \varphi(p) = \lambda\}.$$

One can prove:

1. *For each λ, $0 < \lambda < 1$, there exists a stochastic event φ which can be represented by a finite stochastic automaton, such that the set $E_{\varphi=\lambda}$ is not regular.*

2. *If the stochastic event φ is represented by a finite automaton, then the sets $E_{\varphi=0}$ and $E_{\varphi=1}$ are regular.*

3. *A stochastic event φ on a finite set X which has only a finite number of values is representable by a finite stochastic automaton if and only if for each λ between* 0 *and* 1 *inclusively the set $E_{\varphi=\lambda}$ is regular.*

(Received November 16th, 1965.)

REFERENCES

[1] R. R. Bush and F. Mosteller: Stochastic models for learning. Wiley & Sons, New York — Chapman & Hall, London, 1955.

[2] Ю. А. Шрейдер: Модели обучения и управляющие системы. In: Стохастические модели обучаемости. ИЛ, Москва 1962.

[3] М. Л. Цетлин: Конечные автоматы и моделирование простейших форм поведения. УМН *XXIII* (1963), вып. 4 (112), 3—28.

[4] M. O. Rabin: Probabilistic automata. Information & Control *6* (1963), 230 — 245.

[5] O. Onicescu, S. Guiaşu: Finite abstract random automata. Zeitschrift für Wahrscheinlichkeitstheorie *3* (1965), 279—285.

[6] P. H. Starke: Theorie stochastischer Automaten I, II. Elektronische Informationsverarbeitung und Kybernetik *1* (1965), 5—32, 71—98.

[7] P. H. Starke: Stochastische Ereignisse und Wortmengen. Zeitschrift für Mathematische Logik *12* (1966), 61—68.

[8] P. H. Starke: Stochastische Ereignisse und stochastische Operatoren. Elektronische Informationsverarbeitung und Kybernetik *2* (1966), forthcoming.

[9] K. Nawrotzki: Eine Bemerkung über die Reduktion von stochastischen Automaten. Forthcoming in Elektronische Informationsverarbeitung und Kybernetik.

VÝTAH

Teorie pravděpodobnostních automatů

PETER H. STARKE

Článek, který je shrnutím citovaných prací [6] — [9], je stručným úvodem do teorie pravděpodobnostních automatů, která vznikla během posledních dvou let.

Dipl.-Math. Peter H. Starke, Institut für Mathematische Logik der Humboldt-Universität zu Berlin. Unter den Linden 6, 108 Berlin. DDR.