

**NON-RANDOMIZED MARKOV AND SEMI-MARKOV STRATEGIES  
IN DYNAMIC PROGRAMMING**

*E. A. FAINBERG*

(Translated by *W. U. Sirk*)

**1. Introduction**

In a non-homogeneous controllable Markov model with a total reward criterion, discrete time, infinite horizon and Borel spaces of states and controls, let a certain strategy  $\pi$  and an initial measure  $\mu$  be given. In the paper the following two statements are proved:

(a) (Theorem 3) for any  $K < +\infty$ , there exists a non-randomized Markov strategy  $\varphi$  such that

$$(1) \quad w(\mu, \varphi) \cong \begin{cases} w(\mu, \pi) & \text{if } w(\mu, \pi) < +\infty, \\ K & \text{if } w(\mu, \pi) = +\infty; \end{cases}$$

(b) (Theorem 4) for any measurable function  $K(x) < +\infty$  given on a set of initial states  $X_0$ , there exists a non-randomized semi-Markov strategy  $\varphi'$  such that, for any  $x \in X_0$ ,

$$(2) \quad w(x, \varphi') \cong \begin{cases} w(x, \pi) & \text{if } w(x, \pi) < +\infty, \\ K(x), & \text{if } w(x, \pi) = +\infty. \end{cases}$$

The quantities  $w(\mu, \pi)$  and  $w(x, \pi)$  are the expectations of total reward in the case of the strategy  $\pi$  and initial measure  $\mu$ , and initial state  $x$ , respectively.

Controllable Markov models with Borel state spaces, as well as problems of existence of Markov and semi-Markov strategies in such models which majorize arbitrary strategies, were studied for the first time by Blackwell [1], [2]. These investigations were continued by Strauch [3], where three cases were considered: positive (*P*) and negative (*N*) dynamic programming, as well as dynamic programming with discounting (*D*). For the cases *D* and *N* it was proved, as one of the fundamental results of the investigation [3], Theorem 4.3], that non-randomized Markov strategies  $\varphi$  and semi-Markov strategies  $\varphi'$  such that  $w(\mu, \varphi) \cong w(\mu, \pi)$  and  $w(x, \varphi') \cong w(x, \pi)$  for all initial states  $x$  exist. In all three cases, *D*, *N* and *P*, it was assumed in [3] that  $w(\mu, \pi) < +\infty$  for all  $\mu$  and  $\pi$ , and in view of this the constant  $K$  and the function  $K(x)$  were not considered. For the case *P* (cf. [3], Theorem 4.4), existence of non-randomized Markov strategies  $\varphi$  and semi-Markov strategies  $\varphi'$ , such that  $w(\mu, \varphi) \cong w(\mu, \pi) - \varepsilon$  and  $w(x, \varphi') \cong w(x, \pi) - \varepsilon$  for all initial states  $x$ , was proved for any  $\varepsilon > 0$ . In [3] it

was pointed out that it is not known whether the last result is true for  $\varepsilon = 0$ . (We note that in the formulation of the problem it was assumed in [3] that the initial measure  $\mu$  is concentrated at a single point. The case of an arbitrary initial measure  $\mu$ , for the first time considered by Hinderer [4], does not introduce additional difficulties.)

Homogeneous models were considered in [1]–[3]. The concept of non-homogeneous controllable models arose as a result of the investigations [5]–[7]. In [4], [8] and [9] a considerable part of the investigations [1]–[3] was extended to the case of non-homogeneous models, with a broader class of income functions being investigated in [4] and [9] than in [1]–[3]. For weak conditions the results on existence of a non-randomized Markov strategy in the non-homogeneous case, which majorizes an arbitrary strategy, is presented in [9] Chapt. 5, § 1, Statement II. Also there, for the case  $w(\mu, \pi) < +\infty$ , the question is raised again concerning the existence of a non-randomized Markov strategy  $\varphi$  for which  $w(\mu, \varphi) \cong w(\mu, \pi)$ . A positive answer to this question follows from Theorem 3 of the present paper.

In [4], § 9 and [9], Chapt. 2, § 13, examples are given which show that a situation is possible in which  $w(\mu, \pi) = +\infty$ , but  $w(\mu, \varphi) < +\infty$  for any non-randomized strategy  $\varphi$ . Therefore the conditions  $K < +\infty$  and  $K(x) < +\infty$  are essential in Theorems 3 and 4. The example 3 in [1] shows that in Theorem 4 we cannot assert the existence of a non-randomized Markov strategy that satisfies the inequality (2).

Thus the results of the present paper provide answers to questions in [3], § 4 and [9], Chapt. 5.

The proofs rest on representing a measure in a trajectory space corresponding to an arbitrary (also Markov and semi-Markov) strategy, in terms of measures corresponding to non-randomized (respectively, non-randomized Markov and non-randomized semi-Markov) strategies (Theorem 1). For arbitrary strategies an analogous result was obtained by Krylov [10], Theorem 1, in the case of a finite or denumerable set of states, and by Gikhman and Skorokhod [11], Theorem 1.2, in the case of Borel states and control spaces. This assertion was not considered earlier in connection with Markov and semi-Markov strategies, but it is precisely the assertion of Theorem 1 relating to such strategies which is used in the proof of Theorems 3 and 4. The proof of Theorem 1 presented here is based on Lemma 1.2 in [11] and is carried out according to the same scheme as the proof of Theorem 1 in [10].

The formulation and the proof of Theorem 1 look identical both for controllable Markov models and for models with transition probabilities and income functions depending on the prehistory. Therefore results of the form (1), (2) are first proved for such models and are then transferred to the more particular case of controllable Markov models. Everywhere in the paper we keep to the notation used in [9].

## 2. Basic Definitions

For an arbitrary Borel space  $E$  we shall denote by  $B(E)$  the Borel  $\sigma$ -algebra on  $E$ , and by  $\Pi(E)$  the collection of all probability measures on  $(E, B(E))$ . For

two measurable spaces  $(E_1, \sigma_1)$  and  $(E_2, \sigma_2)$  we denote by  $\Pi(E_2, \sigma_2 | E_1, \sigma_1)$  the set of all transition functions from  $E_1$  into  $E_2$ , i.e., the set of all functions  $P(\Gamma | u_1)$ , where  $\Gamma \in \sigma_2$ ,  $u_1 \in E_1$ , such that  $P(\cdot | u_1)$  is a probability measure on  $E_2$  for any  $u_1$ , and for any  $\Gamma \in \sigma_2$  the function  $P(\Gamma | \cdot)$  is measurable on  $E_1$ . If it is clear about which  $\sigma$ -algebras  $\sigma_1$  and  $\sigma_2$  we are talking, then we shall write  $\Pi(E_2 | E_1)$ .

A *controllable model* (CM) is specified by a set  $\{X, A, j(\cdot), p(\cdot | \cdot), q(\cdot)\}$ . The sets  $X$  and  $A$  are, respectively, *spaces of states and controls*,  $X = \bigcup_{t=1}^{\infty} X_{t-1}$ ,  $A = \bigcup_{t=1}^{\infty} A_t$ , where  $X_{t-1}$  and  $A_t$  are sets of states and controls at the time instants  $t = 1, 2, \dots$ . It is assumed that  $X_{t-1} \cap X_{t'-1} = A_t \cap A_{t'} = \emptyset$  for all  $t \neq t'$ . Further,  $j(a)$  is the mapping of the set  $A$  onto  $X$ , with  $X_{t-1} = j(A_t)$  and for each  $x \in X$  the set of controls in the state  $x$  is  $A(x) = j^{-1}(x)$ . The functions  $p(\cdot | \cdot)$  and  $q(\cdot)$  are, respectively, the *transition functions* and *reward functions*.

The spaces  $X$  and  $A$  are assumed to be Borel spaces, with  $X_{t-1} \in B(X)$ ,  $A_t \in B(A)$  for all  $t = 1, 2, \dots$ . The mapping  $j$  is measurable and may be made uniform, i.e., there exists a measurable mapping  $\psi(x)$  of the space  $X$  into  $A$  such that  $\psi(x) \in A(x)$  for all  $x \in X$ .

We consider the sets  $L = X_0 \times A_1 \times X_1 \times \dots \times A_t \times X_t \times \dots$ ,  $H_{t-1} = X_0 \times A_1 \times X_1 \times \dots \times A_{t-1} \times X_{t-1}$ ,  $\tilde{H}_t = H_{t-1} \times A_t$ , where  $t = 1, 2, \dots$ ,  $H = \bigcup_{t=1}^{\infty} H_{t-1}$ ,  $\tilde{H} = \bigcup_{t=1}^{\infty} \tilde{H}_t$ . Since  $H_{t-1} \cap H_{t'-1} = \tilde{H}_t \cap \tilde{H}_{t'} = \emptyset$  for  $t \neq t'$ , all spaces  $L, H_{t-1}, \tilde{H}_t, H$  and  $\tilde{H}$  are Borel spaces, with the  $\sigma$ -algebras  $B(L), B(H_{t-1}), H_t, B(H)$  and  $B(\tilde{H})$  generated by the  $\sigma$ -algebras  $B(X)$  and  $B(A)$ .

The transition function  $p(\cdot | \cdot)$  is an element of the set  $\Pi(A | \tilde{H})$ ; if  $\tilde{h} \in \tilde{H}_t$  for some  $t = 1, 2, \dots$ , then the measure  $p(\cdot | \tilde{h})$  is concentrated at  $X_t$ . The income function  $q(\cdot)$  is a measurable function on  $\tilde{H}$  which takes values in  $[-\infty, +\infty]$ .

An element  $\pi(\cdot | \cdot) \in \Pi(A | H)$  such that if  $h \in H_{t-1}$  and  $h = x_0 a_1 x_1 \dots a_{t-1} x_{t-1}$  for some  $t = 1, 2, \dots$ , then the measure  $\pi(\cdot | h)$  is concentrated at  $A(x_{t-1})$ , is called a *strategy*  $\pi$ . A necessary and sufficient condition of existence of at least one strategy is the assumption that the mapping  $j$  can be made uniform (cf. [9], Chapt. 3, § 1).

A strategy  $\pi$  is said to be *non-randomized*, if for any prehistory  $h \in H$  the measure  $\pi(\cdot | h)$  is concentrated at a single point. A strategy  $\pi$  is said to be *semi-Markov*, if for any  $z \in X_0$  and  $x \in X \setminus X_0$  there exists a measure  $\pi(\cdot | z, x)$  on  $A$  such that  $\pi(\cdot | h) = \pi(\cdot | x_0, x_{t-1})$  for any  $h = x_0 a_1 \dots a_{t-1} x_{t-1} \in H \setminus H_0$ ,  $t = 2, 3, \dots$ . A semi-Markov strategy  $\pi$  is said to be *Markov*, if for any  $x \in X \setminus X_0$  there exists a measure  $\pi(\cdot | x)$  on  $A$  such that  $\pi(\cdot | x_0, x_{t-1}) = \pi(\cdot | x_{t-1})$  for all  $x_0 \in X_0, x_{t-1} \in X_{t-1}, t = 2, 3, \dots$ . Also *non-randomized semi-Markov and non-randomized Markov strategies* are considered.

We denote by  $\Delta_3^N, \Delta_3, \Delta_2^N, \Delta_2, \Delta_1^N$  and  $\Delta_1$ , respectively, the sets of non-randomized Markov, Markov, non-randomized semi-Markov, semi-Markov, non-randomized and all strategies.

If an initial measure  $\mu \in \Pi(X_0)$  is given, then the strategy  $\pi$  uniquely specifies a probability measure on  $(L, B(L))$  (see [4] or [9]), which we shall denote by  $\mathbf{P}_\mu^\pi$ . We denote by  $\mathbf{E}_\mu^\pi$  the expectation computable from the measure  $\mathbf{P}_\mu^\pi$ . The measure  $\mathbf{P}_\mu^\pi$  induces measures on  $H_{t-1}$  and  $\tilde{H}_t, t = 1, 2, \dots$ , which we also denote by  $\mathbf{P}_\mu^\pi$ .

For an arbitrary number  $g$  we denote  $g^+ = \max(g, 0)$  and  $g^- = -\min(g, 0)$ .

We define the following quantities:

$$w^+(\mu, \pi) = \mathbf{E}_\mu^\pi \sum_{t=1}^\infty q^+(x_0 a_1 x_1 \cdots a_{t-1} x_{t-1} a_t),$$

$$w^-(\mu, \pi) = \mathbf{E}_\mu^\pi \sum_{t=1}^\infty q^-(x_0 a_1 x_1 \cdots a_{t-1} x_{t-1} a_t).$$

If we know that a measure  $\mu \in \Pi(X_0)$  is concentrated at some point  $x$ , then in all notations connected with this measure we shall write  $x$  instead of  $\mu$ .

A CM is said to be  $(\mu, \pi)$ -summable if  $\min\{w^+(\mu, \pi), w^-(\mu, \pi)\} < +\infty$ . If a CM is  $(\mu, \pi)$ -summable, then the value of the *criterion* is estimated by the quantity

$$w(\mu, \pi) = w^+(\mu, \pi) - w^-(\mu, \pi).$$

A CM is said to be  $\mu$ -summable if for any strategy  $\pi$  the CM is  $(\mu, \pi)$ -summable. A CM is said to be weakly  $\pi$ -summable if for any  $x \in X_0$  the CM is  $(x, \pi)$ -summable. If a CM is weakly  $\pi$ -summable for all  $\pi \in \Delta_1$ , then we shall say that the CM is weakly summable. If for any  $\mu \in \Pi(X_0)$  and  $\pi \in A_1$  a CM is  $(\mu, \pi)$ -summable, then such a CM is said to be summable.

A CM is called a *controllable Markov model* (CMM), if there exist a measurable function  $r(a)$  on  $A$  and  $p(\cdot|\cdot) \in \Pi(X|A)$  such that  $r(x_0 a_1 \cdots x_{t-1} a_t) r(a_t)$  and  $p(\cdot|x_0 a_1 \cdots x_{t-1} a_t) = p(\cdot|a_t)$  for all  $\tilde{h} = x_0 a_1 \cdots x_{t-1} a_t \in \tilde{H}$ ,  $t = 1, 2, \dots$ . We note that the definition of a CMM presented here coincides with the definition of a controllable Markov model given in [9].

### 3. Controllable Models

We introduce the notation  $(t = 1, 2, \dots) Y_{t-1}^1 = H_{t-1}$ ,  $Y_{t-1}^3 = X_{t-1}$ ,  $Y_0^2 = X_0$  and  $Y_{t-1}^2 = X_0 \times X_{t-1}$ , if  $t > 1$ . We put  $Y^i = \bigcup_{t=1}^\infty Y_{t-1}^i$ , where  $i = 1, 2, 3$ . Since  $Y_{t-1}^i \cap Y_{t'-1}^i = \emptyset$  for  $t \neq t'$ , the spaces  $Y^i$  are Borel spaces and  $Y_{t-1}^i \in B(Y^i)$  for all  $t = 1, 2, \dots$ .

For  $i = 1, 2, 3$  we denote by  $k_i(y)$  the mapping of  $Y^i$  onto  $X$ , which for each  $t = 0, 1, 2, \dots$  projects the set  $Y_t^i$  onto  $X_t$ , and put  $A(y) = A(k^i(y))$ , where  $y \in Y^i$  (note that  $k_3(y) = y$ ).

According to the definition of Section 2, the strategy  $\pi \in \Delta_i$ , where  $i = 1, 2, 3$ , is an element  $\pi(\cdot|\cdot) \in \Pi(A|Y^i)$  such that  $\pi(A(y)|y) = 1$  for all  $y \in Y^i$ . A non-randomized strategy  $\varphi \in \Delta_i^N$ , where  $i = 1, 2, 3$ , is a measurable mapping  $\varphi(\cdot)$  of the space  $Y^i$  into  $A$  such that  $\varphi(y) \in A(y)$  for all  $y \in Y^i$ .

Let there be given a strategy  $\pi \in \Delta_i$ , where  $i = 1, 2$  or  $3$ . We specify the value  $i$ . By virtue of Lemma 1.2 in [11] there exist  $B([0, 1] \times Y_{t-1}^i)$ -measurable functions  $f_t^i(s, y)$  on  $[0, 1] \times Y_{t-1}^i$  with values in  $A$  such that, for all  $t = 1, 2, \dots$ ,  $y \in Y_{t-1}^i$ ,  $E \in B(A)$ ,

$$(3) \quad m_L(\{s: f_t^i(s, y) \in E\}) = \pi(E|y),$$

where  $m_L$  is the Lebesgue measure on  $[0, 1]$ .

We shall show that  $f_t^i(s, y)$  can be chosen so that  $f_t^i(s, y) \in A(y)$ . Let  $\psi$  be a measurable transformation of  $X$  into  $A$  for which  $\psi(x) \in A(x)$  for all  $x \in X$

(the existence of  $\psi$  was assumed in Section 2). Then  $\tilde{\psi}_i(y) = \psi(k_i(y))$  is a measurable mapping of  $Y^i$  into  $A$ , with  $\tilde{\psi}_i(y) \in A(y)$  for all  $y \in Y^i$ .

For  $t = 1, 2, \dots$ , we define on the sets  $[0, 1] \times Y_{t-1}^i$  the functions

$$(4) \quad \varphi_t^i(s, y) = \begin{cases} f_t^i(s, y) & \text{if } f_t^i(s, y) \in A(y), \\ \tilde{\psi}_i(y) & \text{if } f_t^i(s, y) \notin A(y). \end{cases}$$

We point out that  $\{(s, y): \varphi_t^i(s, y) = f_t^i(s, y)\} = \{(s, y): j(f_t^i(s, y)) = j(\tilde{\psi}_i(y))\}$  and that the measurable function  $j(a)$  assumes values in the Borel space  $X$ . Since any Borel space is either finite or denumerable, or is isomorphic to a straight line, we have

$$\{(s, y): \varphi_t^i(s, y) = f_t^i(s, y)\} \in B([0, 1] \times Y_{t-1}^i).$$

Hence it follows that for each  $t = 1, 2, \dots$ , the function  $\varphi_t^i(s, y)$  is  $B([0, 1] \times Y_{t-1}^i)$ -measurable.

Since  $\pi(A(y)|y) = 1$ , by virtue of (3) and (4),

$$\pi(E|y) = m_L(\{s: \varphi_t^i(s, y) \in E\}),$$

where  $t = 1, 2, \dots$ ,  $y \in Y_{t-1}^i$ ,  $E \in B(A)$ .

We denote by  $\Omega$  the space of sequences  $\omega = \{s_n\}_{n=1}^\infty$ , where  $0 \leq s_n \leq 1$ ,  $n = 1, 2, \dots$ . For fixed  $\pi \in \Delta_i$  each sequence  $\omega = \{s_n\}_{n=1}^\infty \in \Omega$  uniquely specifies a strategy  $\varphi^i[\omega] \in \Delta_i^N$ : if  $y \in Y_{t-1}^i$  for some  $t$ , then

$$(5) \quad \varphi^i[\omega](y) = \varphi_t^i(s_t, y) \in A(y).$$

Let  $B_t$ ,  $t = 1, 2, \dots$ , be minimal  $\sigma$ -algebras on  $\Omega$  generated by the sets  $\{\omega \in \Omega: s_t \in D \in B([0, 1])\}$ . We denote by  $B^i$  the minimal  $\sigma$ -algebra containing  $\{B_t\}_{t=1}^i$ , where  $t = 1, 2, \dots$ , or  $\infty$ , and put  $B^0 = \{\Omega, \emptyset\}$ . Then the functions  $\varphi^i[\omega](y)$  are  $(B^\infty \times B(Y^i))$ -measurable with respect to  $(\omega, y) \in \Omega \times Y^i$ .

We consider on  $(\Omega, B^\infty)$  a measure  $m$  generated by the sequence of independent random quantities  $\{\omega_n\}_{n=1}^\infty$  each having a uniform distribution on  $[0, 1]$ .

**Theorem 1.** *Let there be given a strategy  $\pi \in \Delta_i$ , where  $i = 1, 2$ , or  $3$ , and an arbitrary non-negative (or bounded) measurable function  $f(l)$  on  $L$ . We consider the strategies  $\varphi^i[\omega]$  constructed from  $\pi$  according to (3)–(5).*

A. *The function  $g(\omega, X) = \int_L f(l) \mathbf{P}_x^{\varphi^i[\omega]}(dl)$  is  $(B^\infty \times B(X_0))$ -measurable with respect to the pair of variables  $(\omega, x) \in \Omega \times X_0$ .*

B. *For any  $\mu \in \Pi(X_0)$  the function  $g^*(\omega) = \int_L f(l) \mathbf{P}_\mu^{\varphi^i[\omega]}(dl)$  is  $B^\infty$ -measurable with respect to  $\omega \in \Omega$ .*

C. *The equation*

$$(6) \quad \int_L f(l) \mathbf{P}_\mu^\pi(dl) = \int_\Omega m(d\omega) \int_L f(l) \mathbf{P}_\mu^{\varphi^i[\omega]}(dl)$$

is valid.

PROOF. A. Let  $h = x_0 a_1 x_1 \dots a_{t-1} x_{t-1} \in H_{t-1}$ , where  $t = 1, 2, \dots$ . For  $t \geq 2$ , we use the notation  $\varphi^2[\omega](h) = \varphi^2[\omega](x_0, x_{t-1})$  if  $i = 2$ , but if  $i = 3$ , then we use  $\varphi^3[\omega](h) = \varphi^3[\omega](x_{t-1})$ . Let  $1 \cdot (\cdot)$  be an indicator function,  $A'_t \in B(A_t)$ ,  $X'_t \in B(X_t)$ ,  $t = 1, 2, \dots$ . Then the functions  $\tilde{p}(A'_t | \omega, h) = 1_{\{\varphi^i[\omega](h) \in A'_t\}}(h)$  are  $(B_t \times$

$B(H_{t-1})$ -measurable with respect to  $(\omega, h) \in \Omega \times H_{t-1}$  and, consequently, they are  $(B^\infty \times B(H_{t-1}))$ -measurable. The functions  $\tilde{p}(X'_t | \omega, \tilde{h}) = p(X'_t | \tilde{h})$  are  $(B^\infty \times B(\tilde{H}_t))$ -measurable with respect to  $(\omega, \tilde{h}) \in \Omega \times \tilde{H}_t$ . We introduce the notation  $E_0 = \Omega \times X_0$ ,  $B(E_0) = B^\infty \times B(X_0)$ ,  $E_{2t+1} = A_{t+1}$  and  $E_{2(t+1)} = X_{t+1}$ , where  $t = 0, 1, 2, \dots$ . Since for any  $t = 0, 1, 2, \dots$ , and  $E' \in B(E_{t+1})$ , the function  $\tilde{p}(E' | \cdot)$  defined on  $\prod_{k=0}^t E_k$  is  $B(\prod_{k=0}^t E_k)$ -measurable, statement A follows from [12], Proposition V.1.1. Statement B follows from A.

C. In view of the lemma on multiplicative systems (cf. [9], Supplement 5, Lemma 1 and [13], Chapt. 1, Theorem 20), it is sufficient to prove (6) for all functions  $f(l) = 1_E(l)$  in the case

$$(7) \quad E = D + A_t \times X_t \times A_{t+1} \times \dots,$$

where

$$D = X'_0 \times A'_1 \times X'_1 \times \dots \times A'_{t-1} \times X'_{t-1}, X'_t \in B(X_t),$$

$$A'_t \in B(A_t), t' = 0, 1, 2, \dots, t-1, t'' = 1, 2, \dots, t-1, t = 1, 2, \dots.$$

In this case, (6) has the following form:

$$(8) \quad \mathbf{P}_\mu^\pi(E) = \int_\Omega \mathbf{P}_\mu^{\varphi^{i[\omega]}}(E) m(d\omega).$$

We shall prove (8) by induction. If  $t = 1$ , then  $\mathbf{P}_\mu^\sigma(E) = \mu(X'_0)$  for any strategy  $\sigma$ , and (8) is obvious.

Let, for some  $t$ , the functions  $\mathbf{P}_\mu^{\varphi^{i[\omega]}}(E)$  be  $B^{t-1}$ -measurable for all sets  $E$  of the form (7), and let (8) be valid. Then in view of the theorem on continuation of a measure, the functions  $\mathbf{P}_\mu^{\varphi^{i[\omega]}}(E)$  are  $B^{t-1}$ -measurable and (8) holds for sets  $E$  of the form (7), when  $D \in B(H_{t-1})$ . For arbitrary  $D \in B(H_{t-1})$  and  $A'_t \in B(A_t)$ , we have

$$(9) \quad \begin{aligned} \mathbf{P}_\mu^\pi(D \times A'_t \times X_t \times A_{t+1} \times \dots) &= \mathbf{P}_\mu^\pi(\{\tilde{h} \in D \times A'_t\}) \\ &= \int_D \pi(A'_t | h) \mathbf{P}_\mu^\pi(dh) = \int_D \left[ \int_\Omega \tilde{p}(A'_t | \omega, h) m(d\omega) \right] \int_\Omega \mathbf{P}_\mu^{\varphi^{i[\omega]}}(dh) m(d\omega) \end{aligned}$$

and

$$(10) \quad \begin{aligned} \int_\Omega \mathbf{P}_\mu^{\varphi^{i[\omega]}}(D \times A'_t \times X_t \times A_{t+1} \times X_{t+1} \times \dots) m(d\omega) \\ = \int_\Omega \mathbf{P}_\mu^{\varphi^{i[\omega]}}(\{\tilde{h} \in D \times A'_t\}) m(d\omega) = \int_\Omega m(d\omega) \int_D \tilde{p}(A'_t | \omega, h) \mathbf{P}_\mu^{\varphi^{i[\omega]}}(dh). \end{aligned}$$

We introduce the notation  $(\Omega_1, \mathcal{F}_1) = (H_{t-1}, B(H_{t-1}))$ ,  $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2) = (\Omega, B^\infty, m)$ ,  $\mathcal{F}_2^1 = B^{t-1}$ ,  $\mathcal{F}_2^2 = B$ ,  $y_1 = h$ ,  $y_2 = \omega$ ,  $F(y_1, y_2) = \tilde{p}(A'_t | \omega, h)$ ,  $\mathbf{P}_1(dy_1 | y_2) = \mathbf{P}_\mu^{\varphi^{i[\omega]}}(dh)$ . Then the right sides of the relations (9) and (10) are equal in view of the following statement.

Let there be given a  $\sigma$ -algebra  $(\Omega_1, \mathcal{F}_1)$ , a probability space  $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ ,  $\sigma$ -algebras  $\mathcal{F}_2^1, \mathcal{F}_2^2 \subseteq \mathcal{F}_2$  and  $\mathbf{P}_1(\cdot | \cdot) \in \Pi(\Omega_1, \mathcal{F}_1 | \Omega_2, \mathcal{F}_2^1)$ . Then if the  $\sigma$ -algebras

$\mathcal{F}_2^1$  and  $\mathcal{F}_2^2$  are independent relative to the measure  $\mathbf{P}_2$ , then for any bounded  $(\mathcal{F}_1 \times \mathcal{F}_2^2)$ -measurable function  $F(y_1, y_2)$  on  $\Omega_1 \times \Omega_2$ ,

$$(11) \quad \begin{aligned} & \int_{\Omega_1} \left[ \int_{\Omega_2} F(y_1, y_2) \mathbf{P}_2(dy_2) \right] \int_{\Omega_2} \mathbf{P}_1(dy_1 | y_2) \mathbf{P}_2(dy_2) \\ &= \int_{\Omega_2} \mathbf{P}_2(dy_2) \int_{\Omega_1} F(y_1, y_2) \mathbf{P}_1(dy_1 | y_2). \end{aligned}$$

For the functions  $F(y_1, y_2) = 1_{E_1}(y_1)1_{E_2}(y_2)$ , where  $E_1 \in \mathcal{F}_1$ ,  $E_2 \in \mathcal{F}_2^2$ , equation (11) can easily be verified. Therefore, from the lemma on multiplicative systems (cf. [9] Supplement 5, Lemma 1, and [13] Chapt. 1, Theorem 20) we have (11).

From the equality of the left sides of the relations (9), (10) and the theorem on continuation of a measure, it follows that (8) is valid for all sets

$$(12) \quad E = D' \times X_t \times A_{t+1} \times X_{t+1} \times \cdots,$$

where  $D' \in B(\tilde{H}_t)$ . Since

$$\mathbf{P}_\mu^{\varphi^i[\omega]}(\{\tilde{h} \in D \times A_t'\}) = \int_D \tilde{p}(A_t' | \omega, h) \mathbf{P}_\mu^{\varphi^i[\omega]}(dh),$$

the function  $\tilde{p}(A_t' | \omega, h)$  is  $(B_t \times B(H_{t-1}))$ -measurable with respect to  $(\omega, h)$ , and  $\mathbf{P}_\mu^{\varphi^i[\omega]}(dh) \in \Pi(H_{t-1}, B(H_{t-1}) | \Omega, B^{t-1})$ , where  $h \in H_{t-1}$ ,  $D \in B(H_{t-1})$ ,  $A_t' \in B(A_t)$ , by virtue of Lemma 2 in [9], Supplement 5, the function  $\mathbf{P}_\mu^{\varphi^i[\omega]}(\{\tilde{h} \in D \times A_t'\})$  is  $B^t$ -measurable with respect to  $\omega$ . Consequently, the function  $\mathbf{P}_\mu^{\varphi^i[\omega]}(E)$  is  $B^t$ -measurable for any set of the form (12).

Hence it follows that for any  $D' \in B(\tilde{H}_t)$  and  $X_t' \in B(X_t)$  the functions

$$\mathbf{P}_\mu^{\varphi^i[\omega]}(\{h \in D' \times X_t'\}) = \int_{D'} p(X_t' | \tilde{h}) \mathbf{P}_\mu^{\varphi^i[\omega]}(d\tilde{h})$$

are  $B^t$ -measurable with respect to  $\omega$ . Using Fubini's theorem (see [13], Chapt. II, Theorem 14), we have

$$\begin{aligned} \mathbf{P}_\mu^\pi(D' \times X_t' \times A_{t+1} \times X_{t+1} \times \cdots) &= \int_{D'} p(X_t' | \tilde{h}) \mathbf{P}_\mu^\pi(d\tilde{h}) \\ &= \int_{D'} p(X_t' | \tilde{h}) \int_\Omega \mathbf{P}_\mu^{\varphi^i[\omega]}(d\tilde{h}) m(d\omega) = \int_\Omega m(d\omega) \int_{D'} p(X_t' | \tilde{h}) \mathbf{P}_\mu^{\varphi^i[\omega]}(d\tilde{h}) \\ &= \int_\Omega \mathbf{P}_\mu^{\varphi^i[\omega]}(\{\tilde{h} \in D' \times X_t'\}) m(d\omega) \\ &= \int_\Omega \mathbf{P}_\mu^{\varphi^i[\omega]}(D' \times X_t' \times A_{t+1} \times X_{t+1} \times \cdots) m(d\omega). \end{aligned}$$

The theorem is proved.

**Corollary 1.** *Let there be given an arbitrary strategy  $\pi \in \Delta_i$ , where  $i = 1, 2$  or 3. Then the functions  $w^+(x, \varphi^i[\omega])$  and  $w^-(x, \varphi^i[\omega])$  are  $(B^\infty \times B(X_0))$ -measurable with respect to  $(\omega, x) \in \Omega \times X_0$ , while the functions  $w^+(\mu, \varphi^i[\omega])$  and  $w^-(\mu, \varphi^i[\omega])$  are  $B^\infty$ -measurable with respect to  $\omega \in \Omega$  for all  $\mu \in \Pi(X_0)$ .*

The validity of the corollary follows from Theorem 1. A, B and the definition of the functions  $w^+$  and  $w^-$ .

If a CM is  $(\mu, \pi)$ -summable, then from (6) we have

$$(13) \quad \begin{aligned} w(\mu, \pi) &= w^+(\mu, \pi) - w^-(\mu, \pi) \\ &= \int_{\Omega} w^+(\mu, \varphi^i[\omega])m(d\omega) - \int_{\Omega} w^-(\mu, \varphi^i[\omega])m(d\omega). \end{aligned}$$

We shall consider sets  $\Omega_{\mu} = \{\omega \in \Omega: w^+(\mu, \varphi^i[\omega]) = w^-(\mu, \varphi^i[\omega]) = +\infty\}$ . From the  $(\mu, \pi)$ -summability of CM and (13) it follows that  $m(\Omega_{\mu}) = 0$ . Assuming in Corollary 2 for the sake of definiteness that  $w(\mu, \varphi^i[\omega]) = -\infty$  when  $\omega \in \Omega_{\mu}$ , we obtain the following statement.

**Corollary 2.** *Let there be given a measure  $\mu \in \Pi(X_0)$  and a strategy  $\pi \in \Delta_i$ , where  $i = 1, 2$  or  $3$ , the CM being  $(\mu, \pi)$ -summable. Then*

$$w(\mu, \pi) = \int_D w(\mu, \varphi^i[\omega])m(d\omega).$$

If  $w(\mu, \pi) < +\infty$ , we define the set

$$\Omega(\mu, \pi) = \{\omega \in \Omega \setminus \Omega_{\mu}: w(\mu, \varphi^i[\omega]) \geq w(\mu, \pi)\},$$

but if  $w(\mu, \pi) = +\infty$ , we define for any  $K < +\infty$  the set

$$\Omega(\mu, \pi, K) = \{\omega \in \Omega \setminus \Omega_{\mu}: w(\mu, \varphi^i[\omega]) \geq K\}.$$

From Corollary 2 and the equation  $m(\Omega_{\mu}) = 0$  follows (cf. [9] Lemma 1 § 1.13) the following statement.

**Lemma 1.** *Let there be given  $\mu$  and  $\pi$ , satisfying the conditions of Corollary 2. If  $w(\mu, \pi) < +\infty$ , then  $m(\Omega(\mu, \pi)) > 0$ , but if  $w(\mu, \pi) = +\infty$ , then  $m(\Omega(\mu, \pi, K)) > 0$  for all  $K < +\infty$ .*

**Corollary 3.** *Let there be given a measure  $\mu \in \Pi(X_0)$  and a strategy  $\pi \in \Delta_i$ , where  $i = 1, 2$  or  $3$ , the CM being  $(\mu, \pi)$ -summable. Then for any  $K < +\infty$  there exists a non-randomized strategy  $\varphi \in \Delta_i^N$  such that*

$$w(\mu, \varphi) \geq \begin{cases} w(\mu, \pi) & \text{if } w(\mu, \pi) < +\infty, \\ K & \text{if } w(\mu, \pi) = +\infty. \end{cases}$$

For the verification of the validity of the corollary it is sufficient to choose an arbitrary strategy  $\varphi^i[\omega]$ , where  $\omega$  belongs either to  $\Omega(\mu, \pi)$  or to  $\Omega(\mu, \pi, K)$ , depending on whether  $w(\mu, \pi)$  is or is not finite.

**Corollary 4. A.** *If a CM is  $\mu$ -summable, then  $\sup_{\pi \in \Delta_i} w(\mu, \pi) = \sup_{\varphi \in \Delta_i^N} w(\mu, \varphi)$  for any  $i = 1, 2, 3$ .*

**B.** *If a CM is weakly summable, then  $\sup_{\pi \in \Delta_i} w(x, \pi) = \sup_{\varphi \in \Delta_i^N} w(x, \varphi)$  for any  $i = 1, 2, 3$  and  $x \in X_0$ .*

**C.** *If a CM is summable, then  $\sup_{\pi \in \Delta_i} w(\mu, \pi) = \sup_{\varphi \in \Delta_i^N} w(\mu, \varphi)$  for any  $i = 1, 2, 3$  and  $\mu \in \Pi(X_0)$ .*

We point out that a theorem on decomposition of a randomized strategy, similar to Theorem 1, was presented by Girsanov [10]. For the case  $i = 1$  ( $\pi$  is



an arbitrary strategy, while  $\varphi^i[\omega]$  are arbitrary non-randomized strategies) closely related results have been obtained in [10], Theorem 1 and [11], Theorem 1.2.

For  $i = 1$  the statement of Corollary 3 reinforces Theorems 9.4, 15.2 and 15.4 in [4] and gives the answer to question II, page 66 therein (see Remark 1 below).

**Theorem 2.** *Let there be given a strategy  $\pi \in \Delta_i$ , where  $i = 1$  or 2. Then if a CM is weakly  $\pi$ -summable, for any measurable function  $K(x) < +\infty$  on  $X_0$  there exists a non-randomized strategy  $\varphi \in \Delta_i^N$  such that, for any  $x \in X_0$ ,*

$$w(x, \varphi) \cong \begin{cases} w(x, \pi) & \text{if } w(x, \pi) < +\infty, \\ K(x) & \text{if } w(x, \pi) = +\infty. \end{cases}$$

**PROOF.** Let a value of  $i$  be fixed. We define the function  $w^*(x, \pi) = w(x, \pi)$  if  $w(x, \pi) < +\infty$ , and  $w^*(x, \pi) = K(x)$  if  $w(x, \pi) = +\infty$ . We consider the sets  $Q' = \{(\omega, x) \in \Omega \times X_0: w^+(x, \varphi^i[\omega]) = w^-(x, \varphi^i[\omega]) = +\infty\}$  and  $Q = \{(\omega, x) \in \Omega \times X_0 \setminus Q': w(x, \varphi^i[\omega]) \cong w^*(x, \pi)\}$ . In view of Corollary 1,  $Q \in B^\infty \times B(X_0)$ . The space  $(Q, B(Q))$  is a Borel space.

We consider the mapping  $r$  which projects  $Q$  onto  $X_0$ . We introduce the notation  $Q(x) = r^{-1}(x)$ ,  $x \in X_0$ . For any  $x \in X_0$  we consider the function  $\nu(\Gamma|x) = m(\{\omega: (\omega, x) \in \Gamma\})$ , where  $\Gamma \in B(Q)$ . Then the following statements hold: (i) for any  $\Gamma \in B(Q)$ , the function  $\nu(\Gamma|x)$  is measurable with respect to  $x$ ; (ii) for any  $x \in X_0$ , the measure  $\nu(\cdot|x)$  is concentrated at  $Q(x)$ ; (iii)  $\nu(Q(x)|x) > 0$  for all  $x \in X_0$ .

Property (i) follows from the equation  $\nu(\Gamma|x) = \int_\Omega 1_\Gamma(\omega, x) m(d\omega)$  and measurability of the set  $\Gamma$ . Property (ii) follows directly from the definition of the measure  $\nu(\cdot|x)$ . We introduce the notation

$$\Omega^*(x, \pi) = \begin{cases} \Omega(x, \pi) & \text{if } w(x, \pi) < +\infty, \\ \Omega(x, \pi, K(x)) & \text{if } w(x, \pi) = +\infty. \end{cases}$$

From property (ii) and Lemma 1 we have:  $\nu(Q(x)|x) = m(\Omega^*(x, \pi)) > 0$ . This establishes property (iii).

Since  $Q$  and  $X_0$  are Borel spaces, the mapping  $r$  is measurable, and properties (i)–(iii) are valid, then in view of the Blackwell-Ryll-Nardzhevskii theorem on a measurable choice (cf. [9], Supplement 3, § 2), there exists a measurable mapping  $\xi: X_0 \rightarrow Q$  such that  $\xi(x) \in Q(x)$  for all  $x \in X_0$ .

For  $i = 1, 2$  we denote by  $c_i(y)$  the mapping of  $Y^i$  onto  $X_0$  which for each  $t = 0, 1, 2, \dots$  projects the set  $Y_t^i$  onto  $X_0$ . The projections  $c_i$  are measurable mappings of  $Y^i$  onto  $X_0$ .

We consider a non-randomized strategy  $\varphi \in \Delta_i^N$ :  $\varphi(y) = \varphi^i[\xi(c_i(y))](y)$ , where  $y \in Y^i$ . The mapping  $\varphi(y)$  is measurable in view of  $(B^\infty \times B(Y^i))$ -measurability of the function  $\varphi^i[\omega](y)$  with respect to the pair of variables  $(\omega, y) \in \Omega \times Y^i$  and in view of  $B(Y^i)$ -measurability of the function  $\xi(c_i(y))$ . For each  $x \in X_0$  we have:  $w(x, \varphi) = w(x, \varphi^i[\omega_x]) \cong w^*(x, \pi)$ , where  $\omega_x = \xi(x)$ . The theorem is proved.

The examples in [4], § 9 and [9], Chapt. 1, § 15 show that the condition  $K < +\infty$  in Corollary 3 and the condition  $K(x) < +\infty$  in Theorem 2 are essential.

EXAMPLE 1 (Theorem 2 is not true for  $i = 3$ ). Let  $X_0 = \{b_{-1}, b_0\}$ ,  $X_{t-1} = \{b_{t-1}\}$  for  $t = 2, 3, \dots$ ,  $A_1 = \{d_{-1}, d_0\}$ , where  $d_n = A(b_n)$  for  $n = -1, 0$ ,  $A_2 = \{d_1, d_2\}$ ,  $A_t = \{d_t\}$  for  $t = 3, 4, \dots$ . The reward functions are given by the equations  $q(b_{-1}d_{-1}b_1d_1) = q(b_0d_0b_1d_2) = 1$  and  $q(\tilde{h}) = 0$  for all other prehistories  $\tilde{h}$ . The process deterministically proceeds from  $X_{t-1}$  to  $X_t$ , where  $t = 1, 2, \dots$ . A positive gain is possible only in the state  $b_1$ .

Any two strategies differ only in the state  $b_1$ , since in all remaining states there is only one control. The set  $\Delta_3^N$  is  $\{\varphi_1, \varphi_2\}$ , where  $\varphi^1(b_1) = d_1$ , while  $\varphi^2(b_1) = d_2$ . We consider a strategy  $\pi \in \Delta_3$  for which  $\pi(d_1|b_1) = \pi(d_2|b_1) = 0.5$ . Then  $w(b_{-1}, \pi) = w(b_0, \pi) = 0.5$ , while  $w(b_{-1}, \varphi^1) = w(b_0, \varphi^2) = 1$ ,  $w(b_0, \varphi^1) = w(b_{-1}, \varphi^2) = 0$ .

REMARK 1. Hinderer [4] considered controllable models in which sets of admissible controls depend on the entire prehistory of the process. For these models the results of Section 3 are valid for  $i = 1$  (the proofs in essence remaining the same), but in such models we cannot consider Markov and semi-Markov strategies. We can consider analogously controllable models in which sets of admissible controls depend on initial and current states. For such models the results of Section 3 are valid for  $i = 1, 2$ , but the concept of Markov strategies in them is meaningless.

#### 4. Controllable Markov Models

In Theorem 4.1 of [3] it is proved that in CMM, for any strategy  $\pi$  and measure  $\mu \in \Pi(X_0)$ , there exist Markov strategy  $\pi^* \in \Delta_3$  and a semi-Markov strategy  $\pi^{**} \in \Delta_2$  such that  $w(\mu, \pi) = w(\mu, \pi^*)$  and  $w(x, \pi) = w(x, \pi^{**})$  for all  $x \in X_0$ . In [3] homogeneous CMM were considered; however, the proof of this fact is not altered in going over to non-homogeneous models (cf. [4], Sect. 18, or [9], Chapt. 3, Sect. 8).

The statement of Theorem 3 follows from this result and from Corollary 3 with  $i = 3$ .

**Theorem 3.** *Let there be given an arbitrary strategy  $\pi$ , a measure  $\mu \in \Pi(X_0)$  and a constant  $K < +\infty$ . If a CMM is  $(\mu, \pi)$ -summable, then there exists a non-randomized Markov strategy  $\varphi \in \Delta_3^N$  such that*

$$w(\mu, \varphi) \geq \begin{cases} w(\mu, \pi) & \text{if } w(\mu, \pi) < +\infty, \\ K & \text{if } w(\mu, \pi) = +\infty. \end{cases}$$

**Corollary 5.** A. *If a CMM is  $\mu$ -summable, then*

$$\sup_{\pi \in \Delta_1} w(\mu, \pi) = \sup_{\varphi \in \Delta_3^N} w(\mu, \varphi).$$

B. *If a CMM is weakly summable, then, for all  $x \in X_0$ ,*

$$(14) \quad \sup_{\pi \in \Delta_1} w(x, \pi) = \sup_{\varphi \in \Delta_3^N} w(x, \varphi).$$

C. *If a CMM is summable, then, for all  $\mu \in \Pi(X_0)$ ,*

$$\sup_{\pi \in \Delta_1} w(\mu, \pi) = \sup_{\varphi \in \Delta_3^N} w(\mu, \varphi).$$

Corollary 5B generalizes the result of the paper [14], which is devoted to the proof of equation (14) for homogeneous CMM with a denumerable set of states, under the assumption that the right side of equation (14) is finite for all initial states.

From Theorem 4.1 in [3] and from Theorem 2 for  $i=2$ , we have the following assertion.

**Theorem 4.** *Let there be given an arbitrary strategy  $\pi$  and a measurable function  $K(x) < +\infty$ ,  $x \in X_0$ . If a CMM is weakly  $\pi$ -summable, then there exists a non-randomized semi-Markov strategy  $\varphi \in \Delta_2^N$  for which, for all  $x \in X_0$ ,*

$$w(x, \varphi) \cong \begin{cases} w(x, \pi) & \text{if } w(x, \pi) < +\infty, \\ K(x) & \text{if } w(x, \pi) = +\infty. \end{cases}$$

The example of Blackwell [1], Example 3 (or [9], Chapt. 3, § 9, Example 1) shows that, for non-randomized Markov strategies  $\varphi \in \Delta_3^N$ , Theorem 4 is not true. As was mentioned in Section 1, Theorems 3 and 4 give answers to the questions posed in [3] and [9]; the conditions  $K < +\infty$  and  $K(x) < +\infty$  are essential.

*Received by the editors  
November 28, 1979*

#### REFERENCES

- [1] D. BLACKWELL, *Dynamic programming in problems with decaying action*, Matematika, 11, 4 (1967), pp. 151–160.
- [2] D. BLACKWELL, *Positive dynamic programming*, Matematika, 13, 5 (1969), pp. 103–106.
- [3] R. STRAUCH, *Negative dynamic programming*, Matematika, 13, 5 (1969), pp. 107–127.
- [4] K. HINDERER, *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Springer-Verlag, Berlin-Heidelberg–New York, 1970.
- [5] A. N. SHIRYAEV, *Certain new results in the theory of controllable processes*, Trans. 3rd Prague Conference on Inform. Theory etc., Academia, Prague, 1964, pp. 657–681. (In Russian.)
- [6] A. N. SHIRYAEV, *Certain new results in the theory of controllable processes*, Trans. 4th Prague Conference on Inform. Theory etc., Academia, Prague, 1967, pp. 131–203. (In Russian.)
- [7] E. B. DYNKIN, *Controllable random sequences*, Theory Prob. Appl. 10 (1965), pp. 1–14.
- [8] N. FURUKAWA, *A Markov decision process with non-stationary laws*, Bull. Math. Statist., 13 (1968), pp. 41–52.
- [9] E. B. DYNKIN AND A. A. YUSHKEVICH, *Controllable Markov Processes and Their Applications*, Nauka, Moscow, 1975. (In Russian.)
- [10] N. V. KRYLOV, *Construction of an optimal strategy for a finite controllable chain*, Theory Prob. Appl., 10 (1965), pp. 45–64.
- [11] I. I. GIKHMAN AND A. V. SKOROKHOD, *Controlled Stochastic Processes*, Springer-Verlag, New York, 1979.
- [12] J. NEVEU, *The Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco, 1965.
- [13] P. A. MEYER, *Probability and Potentials*, Blaisdell, Waltham, MA, 1966.
- [14] K. M. VAN HEE, *Markov strategies in dynamic programming*, Math. Oper. Res., 3, 1 (1978), pp. 37–41.