数理解析研究所講究録 1306 巻 2003 年 201-209

Maximizing Order Probabilities on Controlled Markov Chains

長崎県立大学経済学部　植野貴之 (Takayuki Ueno)
Faculty of Economics, Nagasaki Prefectural University
九州大学大学院経済学研究院　岩本誠一 (Seiichi Iwamoto)
Graduate School of Economics, Kyushu University

1 Introduction

It is natural to consider expectation criteria in stochastic decision problems. In particular, both discounted expected value of total reward and average value per stage are well studied criteria in Markov decision processes.

However, in some situation such as in economy of lower growth rate, growth of economy is in itself a preferable criterion to total amount of production such as the gross national production. How can we measure the growth of decision process? We take an order that sequence of earned rewards is nondecreasing in stage direction.

In this paper we consider a probability criterion that the reward is nondecreasing in time—an order probability—. We maximize the order probability on finite-horizon controlled Markov chains. We show that the policy class for maximization depends upon reward function's dependence on today's state, today's decision and tomorrow's state.

In Section 2, we formulate an optimization problem with order probability criterion on finite-stage controlled Markov chains. Section 3 derives a recursive equation for decision process where reward function is independent of today's decision. Section 4 considers process with reward function of today's state, today's decision and tomorrow's state. It is shown that a recursive relation is derived through imbedding method by expanding the original state space.

2 Decision Process with Order Probability

Throughout the paper, the following data is given :

$N \geq 2$ is an integer; the *total number of stages*
$X = \{s_1, s_2, \ldots, s_l\}$ is a finite *state space*
$U = \{a_1, a_2, \ldots, a_k\}$ is a finite *action space*
$r_n : X \times U \to R^1$ is an *n-th reward function* $(0 \leq n \leq N-1)$
$k : X \to R^1$ is a *terminal function*
$p = \{p(\cdot|\cdot,\cdot)\}$ is a *Markov transition law*
: $p(y|x,u) \geq 0 \ \ \forall (x,u,y) \in X \times U \times X, \quad \sum_{y \in X} p(y|x,u) = 1 \ \ \forall (x,u) \in X \times U$

$y \sim p(\cdot\,|x,u)$ denotes that next state y conditioned on state x and action u appears with probability $p(y|x,u)$.

Let an N-stage controlled Markov chain $\{(X_n, U_n)\}$ on finite state space X and finite decision space U be under a Markov transition law p. We maximize the order probability that the reward (random variable) will appear in ascent order

$$r_0(X_0, U_0) \leq r_1(X_1, U_1) \leq \cdots \leq r_{N-1}(X_{N-1}, U_{N-1}) \leq k(X_N).$$

The problem is to how to find an optimal policy which maximizes the order probability $P(r_0 \leq r_1 \leq \cdots \leq r_{N-1} \leq k)$. We focus our attention on policy class where the optimization should be taken.

The *order probability*

$$P(r_0(X_0, U_0) \leq r_1(X_1, U_1) \leq \cdots \leq r_{N-1}(X_{N-1}, U_{N-1}) \leq k(X_N)) \tag{1}$$

depends not only upon initial state but also upon when, where and what the decision maker will choose. We maximize the order probability in Markov class :

$$\mathrm{P}_0(x_0) \qquad \begin{array}{ll} \text{Maximize} & P^{\pi}_{x_0}(r_0 \leq r_1 \leq \cdots \leq r_{N-1} \leq k_N) \\ \text{subject to} & \text{(i)}_n \ X_{n+1} \sim p(\cdot \mid x_n, u_n) \\ & \text{(ii)}_n \ u_n \in U \end{array} \qquad 1 \leq n \leq N$$

Here $P^{\pi}_{x_0}$ is the (discrete) probability measure on the product space X^N induced from the transition law p, a Markov policy $\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\} \in \Pi$, and an initial state $x_0 \in X$. Thus the probability is expressed by the *partial* multiple summation :

$$P^{\pi}_{x_0}(r_0 \leq \cdots \leq k) = \sum\sum\cdots\sum_{(x_1, x_2, \ldots, x_N) \in (*)} p(x_1|x_0, u_0) p(x_2|x_1, u_1) \cdots p(x_N|x_{N-1}, u_{N-1})$$

where the domain $(*)$ in which the partial multiple summation is taken denotes the set of all $(x_1, x_2, \ldots, x_N) \in X \times X \times \cdots \times X$ satisfying

$$r_0(x_0, u_0) \leq r_1(x_1, u_1) \leq \cdots \leq r_{N-1}(x_{N-1}, u_{N-1}) \leq r(x_N).$$

Here we note that the sequence of intermediate decisions $\{u_0, u_1, \cdots, u_{N-1}\}$ is determined through the Markov policy $\pi = \{\pi_0, \ldots, \pi_{N-1}\}$ as follows :

$$u_0 = \pi_0(x_0), \ u_1 = \pi_1(x_1), \ \ldots, \ u_{N-1} = \pi_{N-1}(x_{N-1}).$$

Thus the order probability control problem is written as follows :

$$\mathrm{P}_0(x_0) \qquad \begin{array}{ll} \text{Maximize} & P^{\pi}_{x_0}(r_0 \leq \cdots \leq k) \\ \text{subject to} & \text{(i)}_n, \ \text{(ii)}_n \ 0 \leq n \leq N-1. \end{array}$$

We call this problem the *order problem.* Let $v_0(x_0)$ denote the *maximum value* of problem $\mathrm{P}_0(x_0)$:

$$v_0(x_0) := \operatorname*{Max}_{\pi \in \Pi} P^{\pi}_{x_0}(r_0 \leq \cdots \leq k) \quad x_0 \in X. \tag{2}$$

Then our problem is to find an *optimal* policy π^* in Markov class Π :

$$v_0(x_0) = P^{\pi^*}_{x_0}(r_0 \leq \cdots \leq k) \quad \forall x_0 \in X. \tag{3}$$

3 Subproblems

In this section we derive a recursive equation for process $\mathrm{P}_0(x_0)$. Let $\pi = \{\pi_n, \pi_{n+1}, \cdots, \pi_{N-1}\}$ be a Markov policy for subprocess from n-th stage on. We denote by $\Pi(n)$ the set of all such Markov policies.

Now let us take n $(0 \leq n \leq N-1)$, $x_n \in X$ and a policy $\pi \in \Pi(n)$. We consider the order probability

$$\begin{aligned} & P^{\pi}_{x_n}(r_n \leq r_{n+1} \leq \cdots \leq k) \\ := \; & P^{\pi}_{x_n}(r_n(X_n, U_n) \leq r_{n+1}(X_{n+1}, U_{n+1}) \leq \cdots \leq r_{N-1}(X_{N-1}, U_{N-1}) \leq k(X_N)). \end{aligned}$$

for the process which starts at state $x_n \in X$ and is governed by $\pi \in \Pi(n)$. Formally we set

$$P_{x_N}(k(X_N)) := 1.$$

Then we have the recursive relation:

Lemma 3.1 *For $n \leq N-2$ we have*

$$P^{\pi}_{x_n}(r_n \leq r_{n+1} \leq \cdots \leq k) = \sum_{\substack{* \\ \hookrightarrow \; * \; : \; x_{n+1} \in X \,;\, r_n(x_n, u_n) \leq r_{n+1}(x_{n+1}, u_{n+1})}} P^{\pi'}_{x_{n+1}}(r_{n+1} \leq \cdots \leq k) p(x_{n+1} | x_n, u_n) \tag{4}$$

where $\pi' = \{\pi_{n+1}, \cdots, \pi_{N-1}\}$, $u_m = \pi_m(x_m)$ $m = n, n+1$.

Further for $\pi = \{\pi_{N-1}\}$ we have

$$P^{\pi}_{x_{N-1}}(r_{N-1} \leq k) = \sum_{\substack{* \\ \hookrightarrow \; \star \; : \; x_N \in X \,;\, r_{N-1}(x_{N-1}, u_{N-1}) \leq k(x_N)}} P_{x_N}(k) p(x_N | x_{N-1}, u_{N-1}) \tag{5}$$

where $u_{N-1} = \pi_{N-1}(x_{N-1})$.

We see that Eq. (5) states

$$P^{\pi}_{x_{N-1}}(r_{N-1} \leq k) = \sum_{\star} p(x_N | x_{N-1}, u_{N-1})$$

Now we consider the family of subproblems:

$$\mathrm{P}_n(x_n) \qquad \begin{array}{ll} \text{maximize} & P^{\pi}_{x_n}(r_n \leq r_{n+1} \leq \cdots \leq r_{N-1} \leq k) \\ \text{subject to} & \text{(i)}_{\mathrm{m}} \;\; X_{m+1} \sim p(\cdot \,|\, x_m, u_m) \\ & \text{(ii)}_{\mathrm{m}} \;\; u_m \in U \end{array} \qquad n \leq m \leq N$$

where $0 \leq n \leq N-1$, $x_n \in X$. Let $v_n(x_n)$ denote the maximum value of problem $\mathrm{P}_n(x_n)$:

$$v_n(x_n) := \operatorname*{Max}_{\pi \in \Pi(n)} P^{\pi}_{x_n}(r_n \leq \cdots \leq k) \quad x_n \in X. \tag{6}$$

where we set

$$v_N(x_N) := 1 \quad x_n \in X. \tag{7}$$

Then we have the recursive equation:

Theorem 3.1

$$
\begin{aligned}
v_N(x) &= 1 \qquad x \in X \\
v_{N-1}(x) &= \operatorname*{Max}_{u \in U} \sum_{\substack{* \\ \hookrightarrow \quad * \quad : \quad y \in X\,;\, r_{N-1}(x,u) \le k(y)}} v_N(y)p(y|x,u) \qquad x \in X
\end{aligned} \tag{8}
$$

$$
v_n(x) = \operatorname*{Max}_{u \in U} \sum_{\substack{* \\ \hookrightarrow \quad \star \quad : \quad y \in X\,;\, r_n(x,u) \le r^*_{n+1}(y)}} v_{n+1}(y)p(y|x,u) \qquad x \in X, \quad n = N-2, \cdots, 1, 0. \tag{9}
$$

where $\pi^*_{N-1}, \pi^*_{N-2}, \cdots, \pi^*_1, \pi^*_0$ *are calculated backward ; the first* $\pi^*_{N-1}(x)$ *is a maximizer for (8) and the subsequent* $\pi^*_n(x)$ *is a maximizer for (9). Further,* $r^*_{N-1}, r^*_{N-2}, \cdots, r^*_1$ *are successively defined through* $\pi^*_{N-1}, \pi^*_{N-2}, \cdots, \pi^*_1$ *:*

$$
r^*_n(x) := r_n(x, \pi^*_n(x)) \qquad n = N-1, N-2, \cdots, 1. \tag{10}
$$

3.1 Decision-free Reward System

Now we consider the special case where the reward function $r : X \times U \to R^1$ is independent of decision variable n, u :

$$
r(x,u) = r(x).
$$

Thus the order probability is

$$
P^{\pi}_{x_0}(r_0 \le r_1 \le \cdots \le k) := P^{\pi}_{x_0}(r(X_0) \le r(X_1) \le \cdots \le r(X_{N-1}) \le k(X_N)). \tag{11}
$$

Then we have simplified results.

Corollary 3.1 *For* $n \le N-2$ *we have*

$$
P^{\pi}_{x_n}(r_n \le r_{n+1} \le \cdots \le k) = \sum_{\substack{* \\ \hookrightarrow \quad * \quad : \quad x_{n+1} \in X\,;\, r(x_n) \le r(x_{n+1})}} P^{\pi'}_{x_{n+1}}(r_{n+1} \le \cdots \le k)p(x_{n+1}|x_n, u_n) \tag{12}
$$

where $u_n = \pi_n(x_n)$, $\pi' = \{\pi_{n+1}, \cdots, \pi_{N-1}\}$.

Further for $\pi = \{\pi_{N-1}\}$ *we have*

$$
P^{\pi}_{x_{N-1}}(r_{N-1} \le k) = \sum_{\substack{* \\ \hookrightarrow \quad \star \quad : \quad x_N \in X\,;\, r(x_{N-1}) \le k(x_N)}} P_{x_N}(k)p(x_N|x_{N-1}, u_{N-1}) \tag{13}
$$

where $u_{N-1} = \pi_{N-1}(x_{N-1})$.

Corollary 3.2

$$
\begin{aligned}
v_N(x) &= 1 \qquad x \in X \\
v_{N-1}(x) &= \operatorname*{Max}_{u \in U} \sum_{\substack{y \in X \\ ;r(x) \le k(y)}} v_N(y)p(y|x,u) \qquad x \in X
\end{aligned}
$$

$$
v_n(x) = \operatorname*{Max}_{u \in U} \sum_{\substack{y \in X \\ ;r(x) \le r(y)}} v_{n+1}(y)p(y|x,u) \qquad x \in X, \quad n = N-2, \cdots, 0. \tag{14}
$$

4 Reward Functions Depend on Tomorrow

In this section we treat the reward function which depends not only today but also on tomorrow. We consider both evaluation problem of order probability and optimization problem.

4.1 Evaluation

First we consider a recursive evaluation of order probability. Now let a sequence of reward functions

$$r_n : X\times X \to R^1 \quad (0 \le n \le N-1), \qquad k : X \to R^1$$

be given. We note that the reward functions depend on next state.

$$\mathrm{P}_0(x_0) \qquad \begin{array}{l} \text{Evaluate} \quad P_{x_0}(r_0 \le r_1 \le \cdots \le r_{N-1} \le k) \\ \text{under} \quad (\mathrm{i})_{\mathrm{n}} \quad X_{n+1} \sim p(\cdot\,|x_n) \quad n = 0,\ldots,N-1 \end{array}$$

Thus we evaluate the order probability

$$v_0(x_0) := P_{x_0}(r_0 \le r_1 \le \cdots \le r_{N-1} \le k) \tag{15}$$

under the Markov chain $\{X_n\}_0^N$ with transition probability law $p = \{p(\cdot|\cdot)\}$.

Let us consider the familiy of subproblems $\{\mathrm{P}_n(x_n, x_{n+1})\}$:

$$\mathrm{P}_n(x_n, x_{n+1}) \qquad \begin{array}{l} \text{Evaluate} \quad P_{x_n,x_{n+1}}(r_n \le r_{n+1} \le \cdots \le r_{N-1} \le k) \\ \text{under} \quad (\mathrm{i})_{\mathrm{m}} \quad X_{m+1} \sim p(\cdot\,|x_m) \quad m = n+1,\ldots,N-1 \\ \qquad\qquad (x_n, x_{n+1}) \in X\times X, \quad n = 0,\ldots,N-1. \end{array}$$

Here we note that the evaluated order probability is the conditional probability :

$$\begin{aligned} & P_{x_n,x_{n+1}}(r_n \le r_{n+1} \le \cdots \le r_{N-1} \le k) \\ = \; & P(r_n(X_n, X_{n+1}) \le r_{n+1}(X_{n+1}, X_{n+2}) \le \cdots \le r_{N-1}(X_{N-1}, X_N) \le k(X_N) \\ & \qquad\qquad | \, (X_n, X_{n+1}) = (x_n, x_{n+1})\,) \\ = \; & \sum_{*\,;\, x_{n+2},\cdots,x_N} \sum \cdots \sum p(x_{n+2}|x_{n+1})p(x_{n+3}|x_{n+2}) \cdots p(x_N|x_{N-1}, u_{N-1}) \\ & \hookrightarrow \quad * \;:\; r_n(x_n, x_{n+1}) \le \; r_{n+1}(x_{n+1}, x_{n+2}) \le \; \cdots \le r_{N-1}(x_{N-1}, x_N) \le \; k(x_N). \end{aligned}$$

Let $w_n(x_n, x_{n+1})$ denote the probability of $\mathrm{P}_n(x_n, x_{n+1})$, where

$$w_N(x_N) := 1.$$

Then we have the recursive equation

Lemma 4.1

$$\begin{aligned} w_N(x) &= 1 \qquad\qquad x \in X \\ w_{N-1}(x,y) &= \begin{cases} 1 & \textit{if } r_{N-1}(x,y) \le k(y) \\ 0 & \textit{otherwise} \end{cases} \qquad (x,y) \in X\times X \\ w_n(x,y) &= \sum_{\star} w_{n+1}(y,z)p(z|y) \qquad (x,y) \in X\times X, \quad 0 \le n \le N-1 \qquad (16) \\ & \hookrightarrow \quad \star \;:\; z \in X\,;\, r_n(x,y) \le r_{n+1}(y,z) \end{aligned}$$

The desired probability (15) is given by

$$v_0(x_0) = \sum_{x_1 \in X} w_0(x_0, x_1) p(x_1|x_0)$$

4.2 Optimization

Second we consider a recursive optimization of order probability. Let reward functions

$$r_n : X \times U \times X \to R^1 \quad (0 \leq n \leq N-1), \qquad k : X \to R^1$$

be dependent on well next state as current decision. We optimize the order probability

$$P_{x_0}(r_0 \leq r_1 \leq \cdots \leq r_{N-1} \leq k) \tag{17}$$

under the controlled Markov chain $\{(X_n, U_n)\}$ with transition probability law $p = \{p(\cdot|\cdot,\cdot)\}$. The problem is which class we optimize in and how we can find an optimal policy. The preceding discussion on evaluation enables us to choose an policy class where any policy is a sequence of decision functions :

$$\gamma_n : X \times \Lambda_n \to U \qquad 2 \leq n \leq N-1.$$

Let us introduce the sequence of *yesterday (last) reward sets* $\{\Lambda_n(x_n)\}$ *to current state* x_n :

$$\begin{aligned} \Lambda_0(x_0) &\triangleq \{\lambda_0 \,|\, \lambda_0 = -\infty\} \\ \Lambda_n(x_n) &\triangleq \left\{ \lambda_n \;\middle|\; \begin{array}{c} \lambda_n = r_{n-1}(x_{n-1}, u_{n-1}, x_n) \\ (x_{n-1}, u_{n-1}) \in X \times U \end{array} \right\} \\ & \qquad\qquad x_n \in X, \quad n = 1, \ldots, N. \end{aligned} \tag{18}$$

Further we define *yesterday (last) reward set*

$$\Lambda_n := \bigcup_{x \in X} \Lambda_n(x)$$

and the sequence of *expanded state spaces* $\{Y_n\}$:

$$Y_n := \{y_n = (x_n, \lambda_n) \,|\, x_n \in X,\ \lambda_n \in \Lambda_n(x_n)\} \quad n = 0, \ldots, N.$$

Let us define the corresponding *random variable* $\widetilde{\Lambda}_n$ by

$$\begin{aligned} \widetilde{\Lambda}_0 &\triangleq \lambda_0 \\ \widetilde{\Lambda}_n &\triangleq r_{n-1}(X_{n-1}, U_{n-1}, X_n) \quad n = 1, \ldots, N. \end{aligned}$$

which takes values in Λ_n.

Now we introduce a new controlled Markov chain on the expanded state spaces $\{Y_n\}_0^N$. Here the state variables $\{(X_n; \widetilde{\Lambda}_n)\}$ behave such that the first component $\{X_n\}$ obeys the original Markov transition law p and the second $\{\widetilde{\Lambda}_n\}$ follows the stochastic dynamics $\widetilde{\Lambda}_{n+1} := r_n(x_n, u_n, X_{n+1})$. When the decision-maker chooses a decision $u_n (\in U)$ on $(x_n; \lambda_n)(\in Y_n)$ at n-th stage, the next state random variable $(X_{n+1}; \widetilde{\Lambda}_{n+1})$ will take $(x_{n+1}; \lambda_{n+1})$ with probability

$p(x_{n+1}|x_n, u_n)$ at $(n+1)$-st stage, where $\lambda_{n+1} = r_n(x_n, u_n, x_{n+1})$. Thus this is expressed by a coupled dynamics $(\mathrm{i})_\mathrm{n}$, $(\mathrm{i})'_\mathrm{n}$ $0 \le n \le N-1$.

Thus we maximize the order probability over the new controlled chain on expanded state spaces as follows :

$$\begin{array}{lll} & \text{maximize} & P^{\gamma}_{x_0}(\,\lambda_0 \le r_0 \le r_1 \le \cdots \le r_{N-1} \le k\,) \\ \mathrm{P}_0(x_0;\lambda_0) & \text{subject to} & (\mathrm{i})_\mathrm{n}\ \ X_{n+1} \sim p(\cdot\,|\,x_n, u_n) \\ & & (\mathrm{i})'_\mathrm{n}\ \ \widetilde{\Lambda}_{n+1} = r_n(x_n, u_n, X_{n+1}) \qquad 0 \le n \le N-1 \\ & & (\mathrm{ii})_\mathrm{n}\ \ u_n \in U \end{array}$$

where $\lambda_0 = -\infty$.

We imbed $\mathrm{P}_0(x_0;\lambda_0)$ into the family of subproblems $\{\mathrm{P}_n(x_n;\lambda_n)\}$:

$$\begin{array}{lll} & \text{maximize} & P^{\gamma}_{x_n}(\,\lambda_n \le r_n \le r_{n+1} \le \cdots \le r_{N-1} \le k\,) \\ \mathrm{P}_n(x_n;\lambda_n) & \text{subject to} & (\mathrm{i})_\mathrm{m}\ \ X_{m+1} \sim p(\cdot\,|\,x_m, u_m) \\ & & (\mathrm{i})'_\mathrm{m}\ \ \widetilde{\Lambda}_{m+1} = r_m(x_m, u_m, X_{m+1}) \qquad n \le m \le N-1 \\ & & (\mathrm{ii})_\mathrm{m}\ \ u_m \in U \end{array}$$

where $x_n \in X, \lambda_n \in \Lambda_n(x_n)$ and $0 \le n \le N-1$.

Let $v_n(x_n, \lambda_n)$ denote the maximum value of $\mathrm{P}_n(x_n;\lambda_n)$, where we set

$$v_N(x_N, \lambda_N) := P_{x_N}(\lambda_N \le k(x_N)\,).$$

Then we have the recursive equation:

Theorem 4.1

$$v_N(x;\lambda) = \begin{cases} 1 & \textit{if } \lambda \le k(x) \\ 0 & \textit{otherwise} \end{cases} \qquad x \in X,\ \lambda \in \Lambda_N(x)$$

$$v_{N-1}(x;\lambda) = \operatorname*{Max}_{u\in U} \sum_{*} v_N(\,y;\, r_{N-1}(x,u,y)\,)p(y|x,u) \qquad x \in X,\ \lambda \in \Lambda_{N-1}(x) \tag{19}$$

$$\hookrightarrow \quad * \ : \ y \in X\,;\, \lambda \le r_{N-1}(x,u,y)$$

$$v_n(x;\lambda) = \operatorname*{Max}_{u\in U} \sum_{*} v_{n+1}(\,y;\, r_n(x,u,y)\,)p(y|x,u) \quad x \in X,\ \lambda \in \Lambda_n(x) \quad n = N-2, \cdots, 0 \tag{20}$$

$$\hookrightarrow \quad \star \ : \ y \in X\,;\, \lambda \le r_n(x,u,y)$$

where $\pi^*_{N-1}, \pi^*_{N-2}, \ldots, \pi^*_1, \pi^*_0$ *are calculated backward ; the first* $\pi^*_{N-1}(x;\lambda)$ *is a maximizer for (19) and the subsequent* $\pi^*_n(x;\lambda)$ *is a maximizer for (20).*

References

[1] Altman, E.: Constrained Markov Decision Processes. Chapman & Hall, New York, 1999

[2] Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton, NJ, 1957

[3] Bellman, R.E.: Some Vistas of Modern Mathematics. University of Kentuky Press, Lexington, KY, 1968

[4] Blackwell, D.: Discounted dynamic programming. Ann. Math. Stat. **36**, 226-235 (1965)

[5] Denardo, E.V.: Contraction mappings in the theory underlying dynamic programming. SIAM Review **9**, 165-177 (1968)

[6] Denardo, E.V.: Dynamic Programming: Models and Applications. Prentice-Hall, NJ. 1982

[7] Dynkin, E.B., Yushkevich, A.A.: Controlled Markov Processes. Springer, New York, 1979

[8] Hinderer, K.: Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter. Lectures Notes in Operation Research and Mathematical Systems **33**, Springer, Berlin, 1970

[9] Howard, R.A.: Dynamic Programming and Markov Processes. MIT Press, Cambridge, Mass., 1960

[10] Iwamoto, S.: Theory of Dynamic Program (in Japanese). Kyushu Univ. Press, Fukuoka, 1987

[11] Iwamoto, S.: Associative dynamic programs. J. Math. Anal. Appl. **201**, 195-211 (1996)

[12] Iwamoto, S.: On expected values of Markov statistics. Bull. Informatics and Cybernetics **30**, 1-24 (1998)

[13] Iwamoto, S.: Conditional decision processes with recursive reward function. J. Math. Anal. Appl. **230**, 193-210 (1999)

[14] Iwamoto, S.: "Dynamic Programming", "Principle of Invariant Imbedding" (Japanese) In: Operations Res. Soc. Japan (ed.): Operations Research Dictionary 2000: Basic Ver., pp.229-245, & Terminology Ver. JUSE, Tokyo, 2000

[15] Iwamoto, S., Fujita, T.: Stochastic decision-making in a fuzzy environment. J. Operations Res. Soc. Japan **38**, 467-482 (1995)

[16] Iwamoto, S., Sniedovich, M.: Sequential decision making in fuzzy environment. J. Math. Anal. Appl. **222**, 208-224 (1998)

[17] Iwamoto, S., Tsurusaki, K., Fujita, T.: Conditional decision-making in a fuzzy environment. J. Operations Res. Soc. Japan **42**, 198-218 (1999)

[18] Iwamoto, S., Ueno, T., Fujita, T.: Controlled Markov chains with utility functions. International Workshop on Markov Processes and Controlled Markov Chains, Changsha, Hunan, China, August 22-28, 1999

[19] Karatzas, I., Shreve, S.E.: Methods of Mathematical Finance. Springer, New York, 1998.

[20] Kreps, D.M.: Decision problems with expected utility criteria I. Math. Oper. Res. **2**, 45-53 (1977)

[21] Kreps, D.M.: Decision problems with expected utility criteria II; stationarity. Math. Oper. Res. **2**, 266-274 (1977)

[22] Markovitz, H.: Portfolio selection. J. Finance **8**, 77-91 (1952)

[23] Ozaki, H., Streufert, P.A.: Dynamic programming for non-additive stochastic objects. J. Math. Eco. **25**, 391-442 (1996)

[24] Porteus, E.: An informal look at the principle of optimality. Management Sci. **21**, 1346-1348 (1975)

[25] Porteus, E.: Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs. J. Opt. Theo. Anal. **36**, 419-432 (1982)

[26] Puterman, M.L.: Markov Decision Processes: Stochastic Models. In: Heyman, D.P., Sobel, M.J. (eds.): Handbooks in Operations Research and Management Science Vol. 2, Elsevier, Amsterdam, 1990, Chap. VIII

[27] Puterman, M.L.: Markov Decision Processes: discrete stochastic dynamic programming. Wiley & Sons, New York, 1994

[28] Sniedovich, M.: Dynamic Programming. Marcel Dekker, Inc. NY, 1992

[29] Stokey, N.L., Lucas, R.E..: Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, Mass., 1989

[30] Strauch, R.: Negative dynamic programming. Ann. Math. Stat. **37**, 871-890 (1966)

[31] Streufert, P.A.: Recursive Utility and Dynamic Programming. In: Barberà, S. et al. (eds.): Handbook of Utility Theory Vol. 1, Kluwer, Boston, 1998, Chap. III