No.588

MARKOVIAN DECISION PROCESSES WITH RANDOM OBSERVATIONS

by

Seizo Ikuta

March 1994

MARKOVIAN DECISION PROCESSES WITH RANDOM OBSERVATIONS

Seizo Ikuta
University of Tsukuba

Abstract Decision-making behaviors of not only people in responsible positions in different (managerial, economical, political, and so on) fields of the real world but also us in a daily life can be explained by a sequential stochastic decision process in which decisions are made *after* having obtained certain random observations and examined their content. As the most typical examples of such a sequential stochastic decision process, we can give a problem of buying and/or selling stocks and a job search problem. In the former, random observations are buying and selling prices of stocks of each day, in the latter, the wage of a job that has just been offered. Another examples are a commodity purchasing problem, an asset selling problem, a customer selection problem, an optimal stopping problem, a search problem, and so on. Furthermore, decisions that are daily made by sovereigns of all states can be also regarded as this sort of sequential stochastic decision process in a sense that they make different decisions with taking the stream of everchanging world situations into account and foreseeing their future trends. The purpose of this paper is to propose a basic model by which a class of stochastic decision processes such as being cited above can be well explained and dealt with. Furthermore, it is demonstrated how for the model to be applied to real problems by taking four concrete examples.

1. Introduction

To begin with, we shall notice that decision-making behaviors of not only people in responsible positions in different (managerial, economical, political, and so on) fields of the real world but also us in a daily life can be explained by a sequential stochastic decision process in which decisions are made *after* having obtained certain random observations and examined their content. As the most typical examples of such a sequential stochastic decision process, we can give the following two: a problem of buying and/or selling stocks [31][33] and a job search problem [17][18][32][36]. In the former, random observations are buying and selling prices of stocks of each day, in the latter, the wage of a job that has just been offered. Another examples are a commodity purchasing problem [6][14][20][25], an asset selling problem [13][28], a customer selection problem [10][23], an optimal stopping problem [1][3][9][11][12][22][35], a search problem [2][4][15][16][19][26][27][29], and so on [5][8][7][21][24][28][30][34]. Furthermore, decisions that are daily made by sovereigns of all states can be also regarded as this sort of sequential stochastic decision process in a sense that they make different decisions with taking the stream of everchanging world situations into account and foreseeing their future trends.

The purpose of this paper is to propose a basic model by which a class of stochastic decision processes such as being cited above can be well explained and dealt with. Section 2 that follows provides a strict definition of the basic model, and Section 3 proposes a variation of the basic model, called a disturbance model, where, after making decisions, the process is to be exposed to disturbances. In Section 4 we examine a case that, in the basic model, an action to be taken is defined as not a vector but a scalar. In Section 5 it will be demonstrated how for the model to be well applied to real problems by taking four concrete examples: a

sequential resource allocation problem, two inventory problems, and a customer selection problem.

In the present paper we define the model on some restricted assumptions: discrete time, discrete and finite state and action spaces, discounting factor $\beta < 1$ if a planning horizon is infinite, and so on. In Section 6 we state the possibility of a further generalization of the models.

2. Basic Model

2.1 Description of the Model

Consider the following discrete time stochastic decision process with a finite planning horizon where, for convenience, let points in time be numbered backward from the final point in time of the planning horizon as time 0, time 1, and so on.

A state of the process at each point in time is characterized by two kinds of vectors, called a first-state $\boldsymbol{i} = (i_1, i_2, \cdots, i_I)'$ and a second-state $\boldsymbol{w} = (w_1, w_2, \cdots, w_W)'$ where I and W are both given positive integers. Let the set of all the possible first-states $\boldsymbol{i}$ be designated by $\mathcal{I}$, assumed to be finite. A second-state $\boldsymbol{w}$ is a random sample from a known distribution $F_{\boldsymbol{i}}(\boldsymbol{w})$ with a sample space $\Omega_{\boldsymbol{i}}$ and a finite expectation vector $\boldsymbol{\mu_i} = (\mu_{\boldsymbol{i}1}, \mu_{\boldsymbol{i}2}, \cdots, \mu_{\boldsymbol{i}W})'$. Accordingly, a state space of the process at each point in time can be written by the set $\{(\boldsymbol{i}, \boldsymbol{w}) \mid \boldsymbol{i} \in \mathcal{I},\ \boldsymbol{w} \in \Omega_{\boldsymbol{i}}\}$. Let the set of all the possible actions $\boldsymbol{x} = (x_1, x_2, \cdots, x_{X(\boldsymbol{i})})'$ in first-state $\boldsymbol{i}$ be denoted by $\mathcal{A}(\boldsymbol{i})$, assumed to be finite, where $X(\boldsymbol{i})$ are given positive integers and where each of x_l, $l = 1, 2, \cdots, X(\boldsymbol{i})$, takes one of $n(\boldsymbol{i}, l)$ possible elements, hence the number of elements in $\mathcal{A}(\boldsymbol{i})$ becomes $k(\boldsymbol{i}) = \prod_{l=1}^{\boldsymbol{i}} n(\boldsymbol{i}, l)$ in which $n(\boldsymbol{i}, l)$ are given positive integers. If an action $\boldsymbol{x} \in \mathcal{A}(\boldsymbol{i})$ is taken in state $(\boldsymbol{i}, \boldsymbol{w})$, then an immediate reward $r(\boldsymbol{i}, \boldsymbol{w}, \boldsymbol{x})$ is obtained, assumed to be bounded in $\boldsymbol{i}, \boldsymbol{w}$, and $\boldsymbol{x}$, and the current first-state $\boldsymbol{i}$ changes into $\boldsymbol{j}$ at the next point in time with probability $p(\boldsymbol{j}|\boldsymbol{i}, \boldsymbol{x})$. Finally, let a discount factor be $\beta \in (0, 1]$.

The objective here is to find the optimal decision policy attaining the maximum of the total expected present discounted reward, the expectation of the sum of immediate rewards obtained at each point in time over the finite planning horizon.

2.2 Functional Equation and Optimal Action I

By $u_t(\boldsymbol{i}, \boldsymbol{w})$ we shall denote the maximum of the total expected present discounted reward starting from time t when in state $(\boldsymbol{i}, \boldsymbol{w})$ where $u_0(\boldsymbol{i}, \boldsymbol{w})$ are appropriately defined for the objective function defined for each of concrete decision problems to which the model will be applied; in many cases,

$$u_0(\boldsymbol{i}, \boldsymbol{w}) = \max_{\boldsymbol{x} \in \mathcal{A}(\boldsymbol{i})} r(\boldsymbol{i}, \boldsymbol{w}, \boldsymbol{x}). \tag{2.1}$$

Then, for $t \geq 1$, $u_t(\boldsymbol{i}, \boldsymbol{w})$ can be expressed as

$$u_t(\boldsymbol{i}, \boldsymbol{w}) = \max_{\boldsymbol{x} \in \mathcal{A}(\boldsymbol{i})} \Big\{ r(\boldsymbol{i}, \boldsymbol{w}, \boldsymbol{x}) + \beta \sum_{\boldsymbol{j} \in \mathcal{I}} p(\boldsymbol{j}|\boldsymbol{i}, \boldsymbol{x}) \int_{\boldsymbol{\xi} \in \Omega_{\boldsymbol{j}}} u_{t-1}(\boldsymbol{j}, \boldsymbol{\xi}) dF_{\boldsymbol{j}}(\boldsymbol{\xi}) \Big\} \tag{2.2}$$

where $\boldsymbol{\xi}$ is a second-state that will be observed at the next point in time in first-state $\boldsymbol{j}$. Then, it goes without saying that the optimal action of time t in state $(\boldsymbol{i}, \boldsymbol{w})$ is given by $\boldsymbol{x} = \boldsymbol{x}_t(\boldsymbol{i}, \boldsymbol{w})$ that attains the maximum of the right hand sides of (2.1) and (2.2). Here it is to be noted that (2.1) is equivalent to defining $u_{-1}(\boldsymbol{i}, \boldsymbol{w}) = 0$ in (2.2) for all $\boldsymbol{i}$ and $\boldsymbol{w}$ if we want to define (2.2) for $t \geq 0$ instead of $t \geq 1$. Now, in general, for all t and $\boldsymbol{i}$ define

$$v_t(i) = \int_{w \in \Omega_i} u_t(i, w) dF_i(w). \tag{2.3}$$

Then, (2.2) can be written

$$u_t(i, w) = \max_{x \in \mathcal{A}(i)} \Big\{ r(i, w, x) + \beta \sum_{j \in \mathcal{I}} p(j | i, x) v_{t-1}(j) \Big\}, \tag{2.4}$$

and furthermore we have

$$v_t(i) = \int_{w \in \Omega_i} \max_{x \in \mathcal{A}(i)} \Big\{ r(i, w, x) + \beta \sum_{j \in \mathcal{I}} p(j | i, x) v_{t-1}(j) \Big\} dF_i(w). \tag{2.5}$$

Now, let

$$V_t(i, w, x) = r(i, w, x) + \beta \sum_{j \in \mathcal{I}} p(j | i, x) v_{t-1}(j). \tag{2.6}$$

Then (2.4) and (2.5) can be expressed as, respectively,

$$u_t(i, w) = \max_{x \in \mathcal{A}(i)} V_t(i, w, x), \tag{2.7}$$

$$v_t(i) = \int_{w \in \Omega_i} \max_{x \in \mathcal{A}(i)} V_t(i, w, x) dF_i(w). \tag{2.8}$$

2.3 Optimal Partition

In general, let $\mathcal{A}$ be a set consisting of k X-vectors x_n, $n = 1, 2, \cdots, k$, and let $w = (w_1, w_2, \cdots, w_W)'$, a random variable having a distribution function $F(w)$ with a sample space Ω and a finite expectation vector $\mu = (\mu_1, \mu_2, \cdots, \mu_W)'$. Here k and W are both given nonnegative integers. Then, let us partition the sample space Ω into k mutually exclusive subsets $D(x)$, $x \in \mathcal{A}$, that is, $\bigcup_{x \in \mathcal{A}} D(x) = \Omega$ and $D(x) \bigcap D(y) = \phi$ for any $x \neq y$. Let $\mathcal{D} = \{D(x) \mid x \in \mathcal{A}\}$, called a *partition of Ω on $\mathcal{A}$*, and let a set of all the possible partitions of Ω on $\mathcal{A}$ be denoted by $\Lambda = \{\mathcal{D}\}$, called a *partition space of Ω on $\mathcal{A}$*. Now, for any given function $V(w, x)$, $w \in \Omega$ and $x \in \mathcal{A}$, consider the maximization problem

$$u(w) = \max_{x \in \mathcal{A}} V(w, x), \tag{2.9}$$

and let the expectation of $v(w)$ as to w be denoted by v, i.e.,

$$v = \int_{w \in \Omega} u(w) dF(w) \tag{2.10}$$

$$= \int_{w \in \Omega} \max_{x \in \mathcal{A}} V(w, x) dF(w). \tag{2.11}$$

For any given partition $\mathcal{D} \in \Lambda$, we shall define

$$V(w | \mathcal{D}) = \sum_{x \in \mathcal{A}} V(w, x) I(w \in D(x))^{\dagger} \tag{2.12}$$

and refer to the partition $\mathcal{D}^* = \{D^*(x) \mid x \in \mathcal{A}\}$ attaining the maximum of $V(w | \mathcal{D})$ on Λ for all $w \in \Omega$ as an *optimal partition of Ω on $\mathcal{A}$*, i.e.,

† $I(S)$ is an indicator function, i.e., $I(S) = 1$ if a given statement S is true, or else $I(S) = 0$.

$$V(\boldsymbol{w}|\mathcal{D}^*) = \max_{\mathcal{D}\in\mathcal{A}} V(\boldsymbol{w}|\mathcal{D}). \tag{2.13}$$

Now define, for $n = 1, 2, \cdots, k$,

$$C(\boldsymbol{x}_n) = = \{\boldsymbol{w} \mid V(\boldsymbol{w}, \boldsymbol{x}_n) > V(\boldsymbol{w}, \boldsymbol{x}_a),\ 1 \leq a < n,$$
$$V(\boldsymbol{w}, \boldsymbol{x}_n) \geq V(\boldsymbol{w}, \boldsymbol{x}_b),\ n < b \leq k\}, \tag{2.14}$$
$$\mathcal{C} = \{C(\boldsymbol{x}_n) \mid n = 1, 2, \cdots, k\}. \tag{2.15}$$

Lemma 1 *$\mathcal{C}$ is a partition of Ω on $\mathcal{A}$, and for any given $\boldsymbol{w} \in \Omega$*

$$u(\boldsymbol{w}) = V(\boldsymbol{w}|\mathcal{C}) = \max_{\mathcal{D}\in\mathcal{A}} V(\boldsymbol{w}|\mathcal{D}), \tag{2.16}$$

hence, $\mathcal{C}$ is an optimal partition of Ω on $\mathcal{A}$.

Proof See Appendix B.

The above lemma tells us that the solution of the maximization problem (2.9) for any given $\boldsymbol{w}$ is provided by $\boldsymbol{x} \in \mathcal{A}$ such as $\boldsymbol{w} \in C(\boldsymbol{x})$ because $u(\boldsymbol{w}) = V(\boldsymbol{w}, \boldsymbol{x})$ for such $\boldsymbol{x}$ from (2.16).

Now, from the above lemma we have

$$v = \int_{\boldsymbol{w}\in\Omega} V(\boldsymbol{w}|\mathcal{C})dF(\boldsymbol{w}) \tag{2.17}$$
$$= \sum_{\boldsymbol{x}\in\mathcal{A}} \int_{\boldsymbol{w}\in\Omega} V(\boldsymbol{w}, \boldsymbol{x})I(\boldsymbol{w} \in C(\boldsymbol{x}))dF(\boldsymbol{w}). \tag{2.18}$$
$$= \int_{\boldsymbol{w}\in\Omega} \max_{\mathcal{D}\in\mathcal{A}} V(\boldsymbol{w}|\mathcal{D})dF(\boldsymbol{w}). \tag{2.19}$$

Lemma 2 *The symbols $\int_{\boldsymbol{w}\in\Omega}$ and $\max_{\mathcal{D}\in\mathcal{A}}$ in the right hand side of (2.19) are interchangeable, i.e.,*

$$v = \max_{\mathcal{D}\in\mathcal{A}} \int_{\boldsymbol{w}\in\Omega} V(\boldsymbol{w}|\mathcal{D})dF(\boldsymbol{w}) \tag{2.20}$$
$$= \max_{\mathcal{D}\in\mathcal{A}} \sum_{\boldsymbol{x}\in\mathcal{A}} \int_{\boldsymbol{w}\in\Omega} V(\boldsymbol{w}, \boldsymbol{x})I(\boldsymbol{w} \in D(\boldsymbol{x}))dF(\boldsymbol{w}) \tag{2.21}$$

Proof See Appendix B.

2.4 Functional Equation and Optimal Action II

From (2.21), we can rewrite (2.8) as follows.

$$v_t(i) = \max_{\mathcal{D}(i)\in\mathcal{A}(i)} \sum_{\boldsymbol{x}\in\mathcal{A}(i)} \int_{\boldsymbol{w}\in\Omega_i} V_t(i, \boldsymbol{w}, \boldsymbol{x})I(\boldsymbol{w} \in D(i, \boldsymbol{x}))dF_i(\boldsymbol{w}) \tag{2.22}$$

where $\mathcal{D}(i) = \{D(i, \boldsymbol{x}) \mid \boldsymbol{x} \in A(i)\}$ is a partition of Ω_i on $A(i)$ and $\mathcal{A}(i)$ is a partition space of Ω_i on $A(i)$. Arranging (2.22) by substituting (2.6) into produces

$$v_t(i) = \max_{\mathcal{D}(i)\in\mathcal{A}(i)} \Big\{ \sum_{x\in\mathcal{A}(i)} \int_{w\in\Omega_i} r(i,w,x) I(w\in D(i,x)) dF_i(w)$$

$$+\beta \sum_{j\in\mathcal{I}} \Big(\sum_{x\in\mathcal{A}(i)} p(j|i,x) \int_{w\in\Omega_i} I(w\in D(i,x)) dF_i(w) \Big) v_{t-1}(j) \Big\}, \quad (2.23)$$

which can be rewritten

$$v_t(i) = \max_{\mathcal{D}(i)\in\mathcal{A}(i)} \Big\{ R(i,\mathcal{D}(i)) + \beta \sum_{j\in\mathcal{I}} P(j|i,\mathcal{D}(i)) v_{t-1}(j) \Big\}. \quad (2.24)$$

where

$$R(i,\mathcal{D}(i)) = \sum_{x\in\mathcal{A}(i)} \int_{w\in\Omega_i} r(i,w,x) I(w\in D(i,x)) dF_i(w), \quad (2.25)$$

$$P(j|i,\mathcal{D}(i)) = \sum_{x\in\mathcal{A}(i)} p(j|i,x) \int_{w\in\Omega_i} I(w\in D(i,x)) dF_i(w). \quad (2.26)$$

The above three expressions tell us that the basic model defined in Section 2.1 can be eventually reduced to a standard Markovian decision process with a state space $\mathcal{I}$, an action space $A(i)$, an immediate reward $R(i,\mathcal{D}(i))$, and a state transition probability $P(j|i,\mathcal{D}(i))$. Here it is to be noted that a second-state w seemingly disappears in the system equation (2.24) with being embedded in partition $\mathcal{D}(i)$.

Now, define, for $x_n \in \mathcal{A}(i)$, $n = 1, 2, \cdots, k(i)$,

$$C_t(i,x_n) = \Big\{ w \mid V_t(i,w,x_n) > V_t(i,w,x_a),\ 1 \le a < n,$$

$$V_t(i,w,x_n) \ge V_t(i,w,x_b),\ n < b \le k(i) \Big\}, \quad (2.27)$$

and let

$$\mathcal{C}_t(i) = \Big\{ C_t(i,x_n) \mid n = 1, 2, \cdots, k(i) \Big\}. \quad (2.28)$$

Then, from Lemma 1, $\mathcal{C}_t(i)$ is an optimal partition of Ω_i on $\mathcal{A}(i)$, prescribing the optimal decision rule of time t in first-state i, that is, for any given $w \in \Omega_i$, if $w \in C_t(i,x)$, then take action $x \in \mathcal{A}(i)$. In the case, from (2.18) we have

$$v_t(i) = \sum_{x\in\mathcal{A}(i)} \int_{w\in\Omega_i} V_t(i,w,x) I(w\in C_t(i,x)) dF_i(w), \quad (2.29)$$

which can be expressed as

$$v_t(i) = R(i,\mathcal{C}_t(i)) + \beta \sum_{j\in\mathcal{I}} P(j|i,\mathcal{C}_t(i)) v_{t-1}(j) \quad (2.30)$$

where

$$R(i,\mathcal{C}_t(i)) = \sum_{x\in\mathcal{A}(i)} \int_{w\in\Omega_i} r(i,w,x) I(w\in C_t(i,x)) dF_i(w), \quad (2.31)$$

$$P(j|i,\mathcal{C}_t(i)) = \sum_{x\in\mathcal{A}(i)} p(j|i,x) \int_{w\in\Omega_i} I(w\in C_t(i,x)) dF_i(w). \quad (2.32)$$

2.5 Policy Iteration Algorithm

Here consider the case of infinite planning horizon with discount factor $0 < \beta < 1$. Then let $v(i) = \lim_{t\to\infty} v_t(i)$ where $|v(i)| < \infty$ for all i, and define

$$V(i, w, x) = r(i, w, x) + \beta \sum_{j\in\mathcal{I}} p(j\,|i, x)v(j). \tag{2.33}$$

In the case, a *value determination operation* in Haward's policy iteration algorithm is to solve for a given partition $\mathcal{D}(i) \in \Lambda(i)$, $i \in \mathcal{I}$, the system of equations

$$v(i) = R(i, \mathcal{D}(i)) + \beta \sum_{j\in\mathcal{I}} P(j\,|i, \mathcal{D}(i))v(j), \tag{2.34}$$

and a *policy improvement routine* is to solve for a given $v(i)$, $i \in \mathcal{I}$, the maximization problem

$$\max_{\mathcal{D}(i)\in\Lambda(i)} \Big\{R(i, \mathcal{D}(i)) + \beta \sum_{j\in\mathcal{I}} P(j\,|i, \mathcal{D}(i))v(j)\Big\}. \tag{2.35}$$

2.6 Restrictions of Partition Space

2.6.1 Natural Restriction

In general, the number of all the possible partitions of Ω_i on $\mathcal{A}(i)$ will become enormous, so that finding the optimal partition out of them would require the vast amount of calculations. However, fortunately, if it can be proved in advance that the optimal partition has a specific property $\mathcal{P}(i)$, then the size of partition space $\Lambda(i)$ can be reduced by restricting only partitions with the property. Let the partition space consisting of such partitions be denoted by $\Lambda(i|\mathcal{P}(i)) \subset \Lambda(i)$. We shall call this a *natural restriction* of partition space. Then the constraint "$\mathcal{D}(i) \in \Lambda(i)$" in (2.35) can be replaced with $\mathcal{D}(i) \in \Lambda(i|\mathcal{P}(i))$.

2.6.2 Enforced Restriction

In mathematically analyzing a given concrete sequential stochastic decision process by using the basic model, we very often encounter such a situation that it is quite difficult to strictly prove a property that its optimal partition should or is expected to have or that, even if being proved successfully, it has a too complicated structure to apply it in practice. In the case, if it can be said from an intuitive interpretation that it would have a relatively simple, due property $\mathcal{P}(i)$, optimal or not, then it would not always unreasonable from a practical viewpoint to restrict the optimal partition to one with such a property. We shall refer to such restriction as an *enforced restriction* of partition space. In the case, it is forgivable that we consider Markovian decision process with action space $\Lambda(i|\mathcal{P}(i))$ instead of $\Lambda(i)$ and use the policy iteration algorithm to find the optimal partition on it. Two examples of applications of the enforced restriction will be demonstrated in Section 5.

3. Disturbance Model

Consider the following case. After an action $x \in \mathcal{A}(i)$ has been taken in state (i, w), the process is exposed to a disturbance θ, a random sample from a known distribution $G_i(\theta)$ with a sample space Θ_i, an immediate reward $r(i, w, x, \theta)$ is obtained, and the current

first-state $\boldsymbol{i} \in \mathcal{I}$ changes into a specified $\boldsymbol{j} = j(\boldsymbol{i}, \boldsymbol{x}, \boldsymbol{\theta}) \in \mathcal{I}$ at the next point in time. In the case, $u_t(\boldsymbol{i}, \boldsymbol{w})$ can be expressed as

$$u_t(\boldsymbol{i}, \boldsymbol{w}) = \max_{\boldsymbol{x} \in \mathcal{A}(\boldsymbol{i})} \int_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{i}}} \Big\{ r(\boldsymbol{i}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{\theta}) + \beta \int_{\boldsymbol{\xi} \in \Omega_{\boldsymbol{i}}} u(\boldsymbol{j}, \boldsymbol{\xi}) dF_{\boldsymbol{j}}(\boldsymbol{\xi}) \Big\} dG_{\boldsymbol{i}}(\boldsymbol{\theta}). \tag{3.1}$$

Now define

$$r(\boldsymbol{i}, \boldsymbol{w}, \boldsymbol{x}) = \int_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{i}}} r(\boldsymbol{i}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{\theta}) dG_{\boldsymbol{i}}(\boldsymbol{\theta}). \tag{3.2}$$

If an action $\boldsymbol{x} \in \mathcal{A}(\boldsymbol{i})$ is taken in first-state $\boldsymbol{i}$, then the set of disturbances $\boldsymbol{\theta}$ for which the present first-state $\boldsymbol{i}$ changes into $\boldsymbol{j}$ at the next point in time is given by

$$\mathcal{S}(\boldsymbol{j} \,|\, \boldsymbol{i}, \boldsymbol{x}) = \{\boldsymbol{\theta} \mid \boldsymbol{j} = j(\boldsymbol{i}, \boldsymbol{x}, \boldsymbol{\theta})\}, \tag{3.3}$$

hence, the current first-state $\boldsymbol{i}$ changes into $\boldsymbol{j}$ at the next point in time with probability

$$p(\boldsymbol{j} \,|\, \boldsymbol{i}, \boldsymbol{x}) = \Pr\{\boldsymbol{\theta} \in \mathcal{S}(\boldsymbol{j} \,|\, \boldsymbol{i}, \boldsymbol{x})\}. \tag{3.4}$$

Accordingly, the model with random disturbance can be eventually reduced to the basic model with an immediate reward (3.2) and a state transition probability (3.4). An example of application of this case will be demonstrated in Sections 5.

4. Case of Scalar Action x

Here, we shall consider a case that an action to take at each point in time is defined as not a vector $\boldsymbol{x}$ but a scalar x, so let $\mathcal{A}(\boldsymbol{i}) = \{1, 2, \cdots, k(\boldsymbol{i})\}$. Then, (2.27) can be written

$$\begin{aligned} C_t(\boldsymbol{i}, x) = \{\boldsymbol{w} \mid & V_t(\boldsymbol{i}, \boldsymbol{w}, x) > V_t(\boldsymbol{i}, \boldsymbol{w}, a),\ 1 \le a < x, \\ & V_t(\boldsymbol{i}, \boldsymbol{w}, x) \ge V_t(\boldsymbol{i}, \boldsymbol{w}, b),\ x < b \le k(\boldsymbol{i})\}, \quad 1 \le x \le k(\boldsymbol{i}). \end{aligned} \tag{4.1}$$

In the case, $v_t(\boldsymbol{i})$, $R(\boldsymbol{i}, \mathcal{D}(\boldsymbol{i}))$, $P(\boldsymbol{j} \,|\, \boldsymbol{i}, \mathcal{D}(\boldsymbol{i}))$, $R(\boldsymbol{i}, \mathcal{C}_t(\boldsymbol{i}))$, $P(\boldsymbol{j} \,|\, \boldsymbol{i}, \mathcal{C}_t(\boldsymbol{i}))$, and all other related variables, constants, and expressions are expressed with a replacement of

$$\boldsymbol{x} \implies x \quad \text{and} \quad \sum_{\boldsymbol{x} \in \mathcal{A}(\boldsymbol{i})} \implies \sum_{x=1}^{k(\boldsymbol{i})}$$

Below, in general, for a function $g(x)$ defined on $\mathcal{A}(\boldsymbol{i})$ let $\Delta g(x) = g(x-1) - g(x)$ for $2 \le x \le k(\boldsymbol{i})$ and $\tilde{\Delta} g = g(k(\boldsymbol{i})) - g(1)$. Now, below, for discrete functions we employ the following definition of concavity and convexity.

Concavity and covexity of a discrete function In general, let $g(x)$ be a function of integer variable $x \in [a, b]$ where a and b are both integers. Then, it is said to be concave (convex) in x if $\Delta g(x) = g(x) - g(x-1)$ is nonincreasing (nondecreasing) in x. Then, obviously the definition is equivalent to the following statement: I. $g(\lambda y + (1-\lambda)z) \ge (\le)\ \lambda g(y) + (1-\lambda) g(z)$ for any integers $y, z \in [a, b]$ and any $\lambda \in [0, 1]$ such that $\lambda y + (1-\lambda) z$ is an integer on $[a, b]$. Now, let $\bar{g}(\bar{x})$ be a continuous function that is defined for any real number $\bar{x} \in [a, b]$ by combining successive two points $(x, g(x))$ and $(x+1, g(x+1))$ with a straight line. Then, it is easily seen that, if $g(x)$ is concave (convex) in x in the above sense, so also is $\bar{g}(\bar{x})$ in an ordinary sense, and vise versa. That is, Statement I is equivalent to the following: II. $\bar{g}(\lambda \bar{y} + (1-\lambda)\bar{z}) \ge (\le)\ \lambda \bar{g}(\bar{y}) + (1-\lambda)\bar{g}(\bar{z})$

for any real numbers $\bar{y}, \bar{z} \in [a,b]$ and any real number $\lambda \in [0,1]$. Furthermore, it is also clear that $\max_{x\in[a,b]} g(x) = \max_{\bar{x}\in[a,b]} \bar{g}(\bar{x})$. The above definition will be used also in the next section.

4.1 Concave Case

Let $V_t(\boldsymbol{i}, \boldsymbol{w}, x)$ be concave in x for all t, $\boldsymbol{i}$, and $\boldsymbol{w}$, and define for $2 \le x \le k(\boldsymbol{i})$

$$Y_t(\boldsymbol{i}, x) = \{\boldsymbol{w} \mid \Delta V_t(\boldsymbol{i}, \boldsymbol{w}, x) > 0,\ \boldsymbol{w} \in \Omega_{\boldsymbol{i}}\} \tag{4.2}$$

where

$$\Delta V_t(\boldsymbol{i}, \boldsymbol{w}, x) = \Delta r(\boldsymbol{i}, \boldsymbol{w}, x) + \beta \sum_{\boldsymbol{j}\in\mathcal{I}} \Delta p(\boldsymbol{j}|\boldsymbol{i}, x) v_{t-1}(\boldsymbol{j}). \tag{4.3}$$

Theorem 1 *For all t and $\boldsymbol{i}$, the optimal partition is given by $C_t(\boldsymbol{i})$ that is composed of*

$$C_t(\boldsymbol{i}, x) = \{\boldsymbol{w} \mid \Delta V_t(\boldsymbol{i}, \boldsymbol{w}, x) > 0 \ge \Delta V_t(\boldsymbol{i}, \boldsymbol{w}, x+1)\} \tag{4.4}$$

$$= Y_t(\boldsymbol{i}, x) - Y_t(\boldsymbol{i}, x+1), \qquad 1 \le x \le k(\boldsymbol{i}), \tag{4.5}$$

where $\Delta V_t(\boldsymbol{i}, \boldsymbol{w}, 1) = \infty$ and $\Delta V_t(\boldsymbol{i}, \boldsymbol{w}, k(\boldsymbol{i})+1) = -\infty$, and $Y_t(\boldsymbol{i}, x)$ is nonincreasing in x where $Y_t(k(\boldsymbol{i}), 1) = \Omega_{\boldsymbol{i}}$ and $Y_t(\boldsymbol{i}, k(\boldsymbol{i})+1) = \phi_{\boldsymbol{i}}$. Then we have

$$v_t(\boldsymbol{i}) = \int_{\boldsymbol{w}\in\Omega_{\boldsymbol{i}}} V_t(\boldsymbol{i}, \boldsymbol{w}, 1) dF_{\boldsymbol{i}}(\boldsymbol{w}) + \sum_{x=2}^{k(\boldsymbol{i})} \int_{\boldsymbol{w}\in Y_t(\boldsymbol{i},x)} \Delta V_t(\boldsymbol{i}, \boldsymbol{w}, x) dF_{\boldsymbol{i}}(\boldsymbol{w}). \tag{4.6}$$

Proof See Appendix A.

Policy Iteration Argorithm I

From Theorem 1, as a natural restriction we can restrict the partition space $\Lambda(\boldsymbol{i})$ to $\Lambda(\boldsymbol{i}|\mathcal{P}(\boldsymbol{i}))$ consisting of partitions $\mathcal{D}(\boldsymbol{i}) = \{D(\boldsymbol{i}, x) \mid x \in \mathcal{A}(\boldsymbol{i})\}$, $\boldsymbol{i} \in \mathcal{I}$, that have the property $\mathcal{P}(\boldsymbol{i})$ of $D(\boldsymbol{i}, x) = Z(\boldsymbol{i}, x) - Z(\boldsymbol{i}, x+1)$ with $\Omega_{\boldsymbol{i}} \supseteq Z(\boldsymbol{i}, 1) \supseteq Z(\boldsymbol{i}, 2) \supseteq \cdots \supseteq Z(\boldsymbol{i}, k(\boldsymbol{i})+1)$ where $Z(\boldsymbol{i}, 1) = \Omega_{\boldsymbol{i}}$, and $Z(\boldsymbol{i}, k(\boldsymbol{i})+1) = \phi_{\boldsymbol{i}}$. In the case, (2.25) and (2.26) become, respectively, (See Appendix C)

$$R(\boldsymbol{i}, \mathcal{D}(\boldsymbol{i})) = \int_{\boldsymbol{w}\in\Omega_{\boldsymbol{i}}} r(\boldsymbol{i}, \boldsymbol{w}, 1) dF_{\boldsymbol{i}}(\boldsymbol{w}) + \sum_{x=2}^{k(\boldsymbol{i})} \int_{\boldsymbol{w}\in Z(\boldsymbol{i},x)} \Delta r(\boldsymbol{i}, \boldsymbol{w}, x) dF_{\boldsymbol{i}}(\boldsymbol{w}) \tag{4.7}$$

$$P(\boldsymbol{j}|\boldsymbol{i}, \mathcal{D}(\boldsymbol{i})) = p(\boldsymbol{j}|\boldsymbol{i}, 1) + \sum_{x=2}^{k(\boldsymbol{i})} \Delta p(\boldsymbol{j}|\boldsymbol{i}, x) \int_{\boldsymbol{w}\in Z(\boldsymbol{i},x)} dF_{\boldsymbol{i}}(\boldsymbol{w}). \tag{4.8}$$

In the case, a *value determination operation* is to solve the system of equations (2.34) with (4.7) and (4.8) for a given $\Omega_{\boldsymbol{i}} \supseteq Z(\boldsymbol{i}, 2) \supseteq Z(\boldsymbol{i}, 3) \supseteq \cdots \supseteq Z(\boldsymbol{i}, k(\boldsymbol{i}))$, and a *policy improvement routine* is to find for a given $v(\boldsymbol{i})$, $\boldsymbol{i} \in \mathcal{I}$, the partition $\mathcal{D}(\boldsymbol{i})$ attaining the maximum (2.35) with $\Lambda(\boldsymbol{i}|\mathcal{P}(\boldsymbol{i}))$ instead of $\Lambda(\boldsymbol{i})$. Now, in the same way as the proof of (4.6), the inside of braces in (2.35) can be arranged as follows.

$$\int_{\boldsymbol{w}\in\Omega_{\boldsymbol{i}}} V(\boldsymbol{i}, \boldsymbol{w}, 1) dF(\boldsymbol{w}) + \sum_{x=2}^{k(\boldsymbol{i})} \int_{\boldsymbol{w}\in Z(\boldsymbol{i},x)} \Delta V(\boldsymbol{i}, \boldsymbol{w}, x) dF_{\boldsymbol{i}}(\boldsymbol{w}). \tag{4.9}$$

Hence a *policy improvement routine* can be reduced to the maximization problem

$$G(\boldsymbol{i}) \stackrel{\mathrm{d}}{=} \max_{\Omega_{\boldsymbol{i}} \supseteq Z(\boldsymbol{i},2) \supseteq Z(\boldsymbol{i},3) \supseteq \cdots \supseteq Z(\boldsymbol{i},k(\boldsymbol{i}))} \sum_{x=2}^{k(\boldsymbol{i})} g_x(Z(\boldsymbol{i}, x)), \tag{4.10}$$

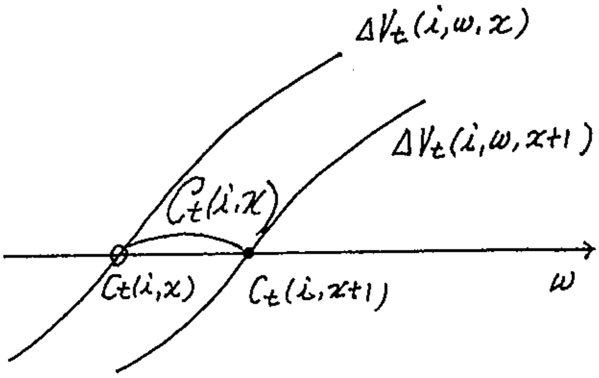

Figure 1

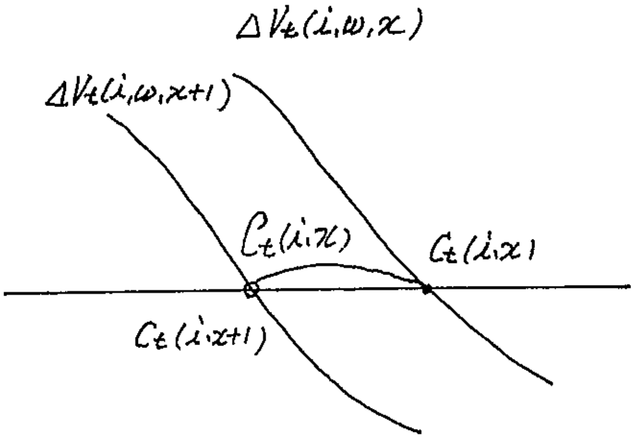

Figure 2

$$g_x(Z(i,x)) = \int_{w\in Z(i,x)} \Delta V(i,w,x)dF_i(w). \tag{4.11}$$

Now define the following maximization problem for any $S \subseteq \Omega_i$ and $x = 2, 3, \cdots, k(i)$:

$$h_x(S) = \max_{\Omega_i \supseteq Z(i,2)\supseteq Z(i,3)\supseteq\cdots\supseteq Z(i,x)\supseteq S} \sum_{l=2}^{x} g_x(Z(i,l)) \tag{4.12}$$

Then, obviously, we have $G(i) = h_{k(i)}(\phi_i)$, and we can easily show that $h_x(S)$ satisfies the following recurrent relation.

$$h_x(S) = \max_{Z(i,x)\supseteq S} \{g_x(Z(i,x)) + h_{x-1}(Z(i,x))\}, \quad 2 \le x \le k(i), \tag{4.13}$$

with $h_1(S) \equiv 0$. Letting the solutions of the maximization problems (4.10) and (4.13) be denoted by, respectively, $Z^*(i,x)$ and $Z(i,x,S)$, $2 \le x \le k(i)$, we have

$$Z^*(i,x) = \begin{cases} Z(i,k(i),\phi_i), & x = k(i), \\ Z(i,x,Z^*(i,x+1)), & x = 2,3,\cdots,k(i)-1. \end{cases} \tag{4.14}$$

■ **Case of Scalar Observation** w

Suppose $\boldsymbol{w}$ is a scalar w, and let the solution of equation $\Delta V_t(i,w,x) = 0$ with unknown w, if exists, be denoted by $c_t(i,x)$. Then, from Theorem 1 and Figures 1 and 2 we have the following corollary.

Corollary 1

(a) ***Case A*** *Assume* $\Delta V_t(i,w,x)$ *is nondecreasing in* w *for all* t, i, *and* x, *and let* $c_t(i,x) = -\infty$ *if* $\Delta V_t(i,w,x) = 0$ *has no solution for a certain* x, *hence* $c_t(i,x') = -\infty$ *for all* $x' \le x$. *Then,* 1. $c_t(i,x)$ *is nondecreasing in* x *where let* $c_t(i,1) = -\infty$ *and* $c_t(i,k(i)+1) = \infty$, 2. $C_t(i,x) = (c_t(i,x), c_t(i,x+1)]$ *and* $Y_t(i,x) = [c_t(i,x), \infty)$, *and* 3. (4.6) *becomes*

$$v_t(i) = \int_{-\infty}^{\infty} V_t(i,w,1)dF_i(w) + \sum_{x=2}^{k(i)} \int_{c_t(i,x)}^{\infty} \Delta V_t(i,w,x)dF_i(w). \tag{4.15}$$

(b) ***Case B*** *Assume* $\Delta V_t(i,w,x)$ *is nonincreasing in* w *for all* t, i, *and* x, *and let* $c_t(i,x) = \infty$ *if* $\Delta V_t(i,w,x) = 0$ *has no solution for a certain* x, *hence* $c_t(i,x') = \infty$ *for all* $x' \le x$. *Then,* 1. $c_t(i,x)$ *is nonincreasing in* x *where let* $c_t(i,1) = \infty$ *and* $c_t(i,k(i)+1) = -\infty$, 2. $C_t(i,x) = [c_t(i,x+1), c_t(i,x))$ *and* $Y_t(i,x) = (-\infty, c_t(i,x)]$, *and* 3. (4.6) *becomes*

$$v_t(i) = \int_{-\infty}^{\infty} V_t(i,w,1)dF_i(w) + \sum_{x=2}^{k(i)} \int_{-\infty}^{c_t(i,x)} \Delta V_t(i,w,x)dF_i(w). \tag{4.16}$$

Policy Iteration Argorithm II

Case A In this case, a *value determination operation* is to solve for a given $y(i,2) \le y(i,3) \le \cdots \le y(i,k(i))$ the system of equations (2.34) with

$$R(i,\mathcal{D}(i)) = \int_{-\infty}^{\infty} r(i,w,1)dF_i(w) + \sum_{x=2}^{k(i)} \int_{y(i,x)}^{\infty} \Delta r(i,w,x)dF_i(w), \tag{4.17}$$

$$P(j|i,\mathcal{D}(i)) = p(j|i,1) + \sum_{x=2}^{k(i)} \Delta p(j|i,x)(1 - F_i(y(i,x))), \tag{4.18}$$

and a *policy improvement routine* is reduced to solving for a given $v(i)$, $i \in \mathcal{I}$, the following recurrent equation

$$h_x(s) = \max_{y(i,x)\leq s} \{g_x(y(i,x)) + h_{x-1}(y(i,x))\}, \quad 2 \leq x \leq k(i). \tag{4.19}$$

Let the solution to the maximization problem of the right hand side of (4.19) be $y(i, x, s)$. Then the solutions to the original problem (4.10), i.e., in the case

$$G(i) \stackrel{\mathrm{d}}{=} \max_{y(i,2)\leq y(i,2)\leq y(i,3)\leq\cdots\leq y(i,k(i))} \sum_{x=2}^{k(i)} g_x(y(i,x)), \tag{4.20}$$

where

$$g_x(y(i,x)) = \int_{y(i,x)}^{\infty} \Delta V(i,w,x)dF_i(w), \tag{4.21}$$

can be given by

$$y^*(i,x) = \begin{cases} y(i,k(i),\infty), & x = k(i), \\ y(i,x,y^*(i,x+1)), & x = 2,3,\cdots,k(i)-1. \end{cases} \tag{4.22}$$

Case B In this case, a *value determination operation* is to solve for a given $y(i,2) \geq y(i,3) \geq \cdots \geq y(i,k(i))$ the system of equations (2.34) with

$$R(i,\mathcal{D}(i)) = \int_{-\infty}^{\infty} r(i,w,1)dF_i(w) + \sum_{x=2}^{k(i)} \int_{-\infty}^{y(i,x)} \Delta r(i,w,x)dF_i(w), \tag{4.23}$$

$$P(j|i,\mathcal{D}(i)) = p(j|i,1) + \sum_{x=2}^{k(i)} \Delta p(j|i,x)F_i(y(i,x)), \tag{4.24}$$

and a *policy improvement routine* is reduced to solving for a given $v(i)$, $i \in \mathcal{I}$, the following recurrent equation

$$h_x(s) = \max_{y(i,x)\geq s} \{g_x(y(i,x)) + h_{x-1}(y(i,x))\}, \quad 2 \leq x \leq k(i). \tag{4.25}$$

Let the solution to the maximization problem of the right hand side of (4.10) be $y(i, x, s)$. Then the solutions of the original problem (4.10), i.e., in the case

$$G(i) \stackrel{\mathrm{d}}{=} \max_{y(i,2)\leq y(i,2)\leq y(i,3)\leq\cdots\leq y(i,k(i))} \sum_{x=2}^{k(i)} g_x(y(i,x)), \tag{4.26}$$

where

$$g_x(y(i,x)) = \int_{y(i,x)}^{\infty} \Delta V(i,w,x)dF_i(w). \tag{4.27}$$

can be given by

$$y^*(i,x) = \begin{cases} y(i,k(i),-\infty), & x = k(i), \\ y(i,x,y^*(i,x+1)), & x = 2,3,\cdots,k(i)-1. \end{cases} \tag{4.28}$$

■ Separable case

Suppose $\boldsymbol{w}$ is a scalar w and $r(i,w,x)$ is expressed as follows.

$$r(i,w,x) = r(i,x)w + e(i,x) \tag{4.29}$$

where $r(i,x)$ is assumed to be either *strictly increasing* or *strictly decreasing* in x for all i. Then $V_t(i,w,x)$ can be expressed as

$$V_t(i,w,x) = r(i,x)w + z_t(i,x) \tag{4.30}$$

where

$$z_t(i,x) = e(i,x) + \beta \sum_{j\in\mathcal{I}} p(j|i,x)v_{t-1}(j). \tag{4.31}$$

Then

$$\Delta V_t(i,w,x) = \Delta r(i,x)w + \Delta z_t(i,x) = \Delta r(i,x)(w - c_t(i,x)), \tag{4.32}$$

where

$$c_t(i,x) = -\Delta z_t(i,x)/\Delta r(i,x), \tag{4.33}$$

which is the solution of $\Delta V_t(i,w,x) = 0$. Now we shall define the following two functions:

$$T_i(y,z) = \int_y^\infty (w-z)dF_i(w), \qquad T_i(z) = T_i(z,z), \tag{4.34}$$

$$S_i(y,z) = \int_{-\infty}^y (w-z)dF_i(w), \qquad S_i(z) = S_i(z,z). \tag{4.35}$$

Then, from Corollary 1 we have:

Case A If $r(i,x)$ is strictly increasing in x, then we have

$$v_t(i) = r(i,1)\mu_i + z_t(i,1) + \sum_{x=2}^{k(i)} \Delta r(i,x)T_i(c_t(i,x)). \tag{4.36}$$

Case B If $r(i,x)$ is strictly decreasing in x, then we have

$$v_t(i) = r(i,1)\mu_i + z_t(i,1) + \sum_{x=2}^{k(i)} \Delta r(i,x)S_i(c_t(i,x)). \tag{4.37}$$

where μ_i is the expectation of the scalar random variable w.

Policy Iteration Algorithm III

Now we have

$$\Delta r(i,w,x) = \Delta r(i,x)(w - h(i,x)) \tag{4.38}$$

where

$$h(i,x) = -\Delta e(i,x)/\Delta r(i,x). \tag{4.39}$$

Case A A *value determination operation* is to solve for a given $y(i,2) \le y(i,3) \le \cdots \le y(i,k(i))$ the system of equations (2.34) with (4.18) and

$$R(i,\mathcal{D}(i)) = r(i,1)\mu_i + e(i,1) + \sum_{x=2}^{k(i)} \Delta r(i,x)T_i(y(i,x),h(i,x)), \tag{4.40}$$

and a *policy improvement routine* is reduced to solving (4.19) and (4.20) where

$$g_x(y(i,x)) = \Delta r(i,x)T_i(y(i,x),c(i,x)). \tag{4.41}$$

Case B A *value determination operation* is to solve for a certain given $y(i,2) \ge y(i,3) \ge \cdots \ge y(i,k(i))$ the system of equations (2.34) with (4.24) and

$$R(i,\mathcal{D}(i)) = r(i,1)\mu_i + e(i,1) + \sum_{x=2}^{k(i)} \Delta r(i,x)S(y(i,x),h(i,x)), \tag{4.42}$$

and a *policy improvement routine* is reduced to solving (4.25) and (4.26) where

$$g_x(y(\boldsymbol{i}, x)) = \Delta r(\boldsymbol{i}, x) S_{\boldsymbol{i}}(y(\boldsymbol{i}, x), c(\boldsymbol{i}, x)). \tag{4.43}$$

■ **Additive Case**

Consider the case that a first-state consists of two entities, $\boldsymbol{i} \in \mathcal{I}$ and $l \in \mathcal{L} = \{1, 2, \cdots, L\}$ with a given nonnegative integer L, and a second-state is a l-vector $\boldsymbol{w}_l = (w_{l1}, w_{l2}, \cdots, w_{ll})'$, depending on l, which is a random variable having a known distribution function $F_{\boldsymbol{i}l}(\boldsymbol{w}_l)$ with a finite expectation vector $\boldsymbol{\mu}_{\boldsymbol{i}l} = (\mu_{\boldsymbol{i}l1}, \mu_{\boldsymbol{i}l2}, \cdots, \mu_{\boldsymbol{i}ll})'$ and a sample space $\Omega_{\boldsymbol{i}l}$. Let an action space in first-state $(\boldsymbol{i}, l)$ be defined by $\mathcal{A}(\boldsymbol{i}, l) = \{0, 1, \cdots, k(\boldsymbol{i}, l)\}$ with $0 \le k(\boldsymbol{i}, l) \le l$. Let $p(\boldsymbol{j}|\boldsymbol{i}, x)$ be the probability that a current $\boldsymbol{i} \in \mathcal{I}$ changes into $\boldsymbol{j} \in \mathcal{I}$ at the next point in time, provided that an action $x \in \mathcal{A}(\boldsymbol{i}, l)$ is taken, and let $q(m|l)$ be the probability that a current $l \in \mathcal{L}$ changes into $m \in \mathcal{L}$ at the next point in time. Therefore, the probability that a current first-state $(\boldsymbol{i}, l)$ changes into $(\boldsymbol{j}, m)$ in the next point in time if an action $x \in \mathcal{A}(\boldsymbol{i}, l)$ is taken becomes

$$p(\boldsymbol{j}, m|\boldsymbol{i}, l, x) = p(\boldsymbol{j}|\boldsymbol{i}, x) q(m|l). \tag{4.44}$$

If an action $x \in \mathcal{A}(\boldsymbol{i}, l)$ is taken in a state $(\boldsymbol{i}, l, \boldsymbol{w}_l)$, then an immediate reward

$$r(\boldsymbol{i}, l, \boldsymbol{w}_l, x) = \sum_{n=0}^{x} w_{ln} \tag{4.45}$$

is obtained where $w_{l0} \equiv 0$, a dummy. In the case, (2.6) can be written as follows.

$$V_t(\boldsymbol{i}, l, \boldsymbol{w}_l, x) = \sum_{n=0}^{x} w_{ln} + z_t(\boldsymbol{i}, l, x), \quad 0 \le x \le k(\boldsymbol{i}, l), \tag{4.46}$$

where

$$z_t(\boldsymbol{i}, l, x) = \beta \sum_{\boldsymbol{j} \in \mathcal{I}} \sum_{m=0}^{L} p(\boldsymbol{j}|\boldsymbol{i}, x) q(m|l) v_{t-1}(\boldsymbol{j}, m). \tag{4.47}$$

Then, we have for $1 \le x \le k(\boldsymbol{i}, l)$

$$\Delta V_t(\boldsymbol{i}, l, \boldsymbol{w}_l, x) = w_{lx} - c_t(\boldsymbol{i}, l, x) \tag{4.48}$$

where

$$c_t(\boldsymbol{i}, l, x) = -\Delta z_t(\boldsymbol{i}, l, x). \tag{4.49}$$

Furthermore from (4.48) we have

$$Y_t(\boldsymbol{i}, l, x) = \{\boldsymbol{w}_l \mid w_{lx} \ge c_t(\boldsymbol{i}, l, x),\ \boldsymbol{w}_l \in \Omega_{\boldsymbol{i}l}\} \tag{4.50}$$

for $1 \le x \le k(\boldsymbol{i}, l)$. Let $Y_t(\boldsymbol{i}, l, 0) = \Omega_{\boldsymbol{i}l}$ and $Y_t(\boldsymbol{i}, l, k(\boldsymbol{i}, l) + 1) = \phi_{\boldsymbol{i}l}$. Now, let $F_{\boldsymbol{i}lx}(w_{lx})$ be a marginal distribution function of $F_{\boldsymbol{i}l}(\boldsymbol{w}_l)$ with respect to w_{lx}, and define

$$T_{\boldsymbol{i}lx}(y, z) = \int_y^{\infty} (w_{lx} - z) dF_{\boldsymbol{i}lx}(w_{lx}), \qquad T_{\boldsymbol{i}lx}(z) = T_{\boldsymbol{i}lx}(z, z). \tag{4.51}$$

Then, from Theorem 1 we have immediately the following corollary.

Corollary 2 *Assume that $V_t(\boldsymbol{i}, l, w_l, x)$ is concave in x for all $t, \boldsymbol{i}, l$, and w_l. Then, the optimal partition of $\Omega_{\boldsymbol{i}l}$ on $\mathcal{A}(\boldsymbol{i}, l)$ is given by $\mathcal{C}(\boldsymbol{i}, l) = \{C(\boldsymbol{i}, l, x) \mid 0 \le x \le k(\boldsymbol{i}, l)\}$ that is composed of*

$$C_t(i,l,x) = \{w_l \mid c_t(i,l,x) < w_{lx},\ w_{l,x+1} \le c_t(i,l,x+1)\} \tag{4.52}$$

where $c_t(i,l,x)$ *is nondecreasing in* x *with* $c_t(i,l,0) = -\infty$, $c_t(i,l,k(i,l)+1) = \infty$, *and* $w_{l,k(i,l)+1} = -\infty$. *Then we have*

$$v_t(i,l) = z_t(i,l,0) + \sum_{x=1}^{k(i,l)} T_{ilx}(c_t(i,l,x)). \tag{4.53}$$

Policy Iteration Algorithm IV

In this case, it is sufficient to consider only partitions $\mathcal{D}(i,l)$ that is composed of

$$D(i,l,x) = \{w_l \mid y(i,l,x) < w_x,\ w_{x+1} \le y(i,l,x+1)\} \tag{4.54}$$

where $y(i,l,1) \le y(i,l,2) \le \cdots \le y(i,l,k(i,l))$, $y(i,l,0) = -\infty$, and $y(i,l,k(i,l)+1) = \infty$. Then, a *value determination operation* is to solve for a certain given $y(i,l,1) \le y(i,l,2) \le \cdots \le y(i,l,k(i,l))$ the system of equations

$$v(i,l) = R(i,l,\mathcal{D}(i,l)) + \beta \sum_{j\in\mathcal{I}} \sum_{m=0}^{\infty} P(i,m|j,l,\mathcal{D}(i,l))v(j,k) \tag{4.55}$$

where from (4.7) and (4.8)

$$R(i,l,\mathcal{D}(i,l)) = \sum_{x=1}^{k(i,l)} \int_{y(i,l,x)}^{\infty} w_{lx} dF_{ilx}(w_{lx}) \tag{4.56}$$

$$P(j,m|i,l,\mathcal{D}(i,l)) = \Big(p(j|i,0) + \sum_{x=1}^{k(i,l)} \Delta p(j|i,x) F_{ilx}(y(i,l,x))\Big) q(m|l), \tag{4.57}$$

and a *policy improvement routine* is reduced to solving for a given $v(i,l)$, $i \in \mathcal{I}$ and $l \in \mathcal{L}$, the maximization problem

$$\max_{y(i,l,1)\le y(i,l,2)\le\cdots\le y(i,l,k(i,l))} \sum_{x=1}^{k(i,l)} T_{ilx}(y(i,l,x), c(i,l,x)). \tag{4.58}$$

5.2 Convex Case

Let $V_t(i,w,x)$ is concave in x for all t, i, and w, and define

$$Y_t(i) = \{w \mid \tilde{\Delta} V_t(i,w) > 0\} \tag{4.59}$$

where

$$\tilde{\Delta} V_t(i,w) = \tilde{\Delta} r(i,w) + \beta \sum_{j\in\mathcal{J}} \tilde{\Delta} p(j|i) v_{t-1}(j). \tag{4.60}$$

Theorem 2 *For all* t *and* i, *the optimal partition is given by* $C_t(i)$ *that is composed of*

$$C_t(i,x) = \begin{cases} \{w \mid \tilde{\Delta} V_t(i,w) \le 0\} & x = 1, \\ \phi_i & x = 2,3,\cdots,k(i), \\ \{w \mid \tilde{\Delta} V_t(i,w) > 0\} & x = k(i), \end{cases} \tag{4.61}$$

and we have

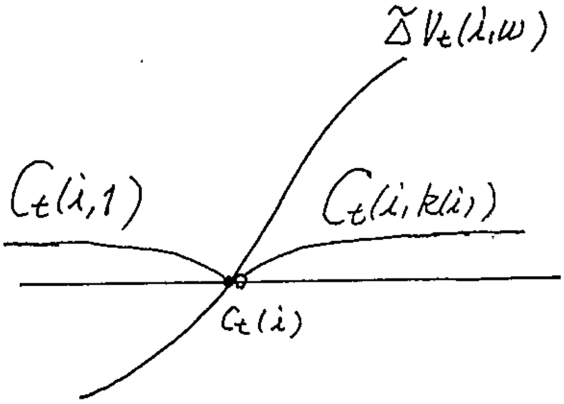

Figure 3

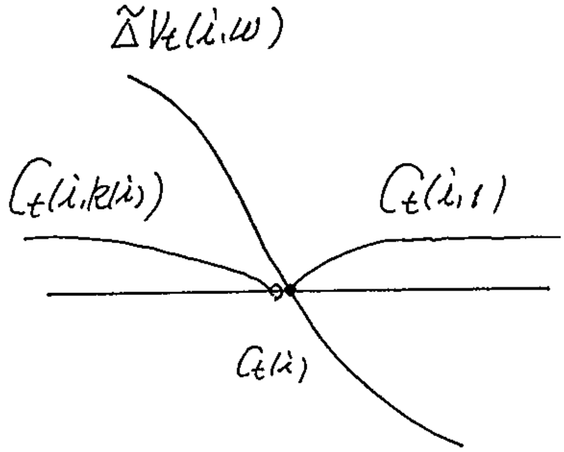

Figure 4

$$v_t(i) = \int_{w \in \Omega_i} V_t(i, w, 1) dF(w) + \int_{w \in Y_t(i)} \tilde{\Delta} V_t(i, w) dF_i(w). \tag{4.62}$$

Proof See Appendix A.

Policy Iteration Argorithm V

From Theorem 2, as a natural restriction, we can restrict the partition space $\Lambda(i)$ to $\Lambda(i|\mathcal{P}(i))$ consisting of partitions $\mathcal{D}(i) = \{D(i,x) \mid x \in \mathcal{A}(i)\}$, $i \in \mathcal{I}$, that have the property $\mathcal{P}(i)$ of $D(i,1) = \Omega_i - Z(i)$, $D(i,x) = \phi_i$ for $1 < x < k(i)$, and $D(i,k(i)) = Z(i)$ where $Z(i) \in \Omega_i$. In the case, (2.25) and (2.26) become (See Appendix C), respectively,

$$R(i, \mathcal{D}(i)) = \int_{w \in \Omega_i} r(i, w, 1) dF(w) + \int_{w \in Z(i)} \tilde{\Delta} r(i, w) dF_i(w), \tag{4.63}$$

$$P(j|i, \mathcal{D}(i)) = p(j|i, 1) + \tilde{\Delta} p(j|i) \int_{w \in Z(i)} dF_i(w). \tag{4.64}$$

In the case, a *value determination operation* is to solve the system of equations (2.34) with (4.63) and (4.64) for a given $Z(i) \in \Omega_i$, $i \in \mathcal{I}$, and a *policy improvement routine* is to find for a given $v(i)$, $i \in \mathcal{I}$, the partition $\mathcal{D}(i)$ attaining the maximum (2.35) with $\Lambda(i|\mathcal{P}(i))$ instead of $\Lambda(i)$. Now, in the same way as the proof of (4.62), the inside of braces in (2.35) can be arranged as follows.

$$\int_{w \in \Omega_i} V(i, w, 1) dF(w) + \int_{w \in Z(i)} \tilde{\Delta} V(i, w) dF_i(w). \tag{4.65}$$

Consequently, a *policy improvement routine* can be reduced to the maximization problem

$$g(Z(i)) = \max_{Z(i) \in \Omega_i} \int_{w \in Z(i)} \Delta V(i, w) dF_i(w). \tag{4.66}$$

■ **Case of Scalar Observation** w

Suppose $\boldsymbol{w}$ is a scalar w, and let the solution of equation $\Delta V_t(i, w) = 0$ with unknown w, if exists, be denoted by $c_t(i)$. Then, from Theorem 2 and Figures 3 and 4 we have the following corollary.

Corollary 3

(a) ***Case A*** *Assume $\tilde{\Delta} V_t(i, w)$ is nondecreasing in w for all t and i, and let $c_t(i) = -\infty$ if $\tilde{\Delta} V_t(i, w, x) = 0$ has no solution. Then, $C_t(i,1) = (-\infty, c_t(i)]$, $C_t(i, k(i)) = (c_t(i), \infty)$, and $Y_t(i) = (c_t(i), \infty)$, and (4.62) becomes*

$$v_t(i) = \int_{-\infty}^{\infty} V_t(i, w, 1) dF_i(w) + \int_{c_t(i)}^{\infty} \tilde{\Delta} V_t(i, w) dF_i(w). \tag{4.67}$$

(b) ***Case B*** *Assume $\tilde{\Delta} V_t(i, w)$ is nonincreasing in w for all t and i, and let $c_t(i) = \infty$ if $\tilde{\Delta} V_t(i, w) = 0$ has no solution. Then, $C_t(i,1) = [c_t(i), \infty)$, $C_t(i, k(i)) = (-\infty, c_t(i))$, and $Y_t(i) = (-\infty, c_t(i))$, and (4.62) becomes*

$$v_t(i) = \int_{-\infty}^{\infty} V_t(i, w, 1) dF_i(w) + \int_{-\infty}^{c_t(i)} \tilde{\Delta} V_t(i, w) dF_i(w). \tag{4.68}$$

Policy Iteration Argorithm VI

Case A In the case, a *value determination operation* is to solve for a certain given $y(i)$ the system of equations (2.34) with

$$R(i,\mathcal{D}(i)) = \int_{-\infty}^{\infty} r(i,w,1)dF_i(w) + \int_{y(i)}^{\infty} \tilde{\Delta}r(i,w)dF_i(w), \tag{4.69}$$

$$P(j|i,\mathcal{D}(i)) = p(j|i,1) + \tilde{\Delta}p(j|i)(1 - F_i(y(i))), \tag{4.70}$$

and a *policy improvement routine* is to solve for a certain given $v(i)$, $i \in \mathcal{I}$, the maximization problem

$$g(i) = \max_{y(i)} \int_{y(i)}^{\infty} \tilde{\Delta}V(i,w)dF_i(w). \tag{4.71}$$

Case B In the case, a *value determination operation* is to solve for a certain given $y(i)$ the system of equations (2.34) with

$$R(i,\mathcal{D}(i)) = \int_{-\infty}^{\infty} r(i,w,1)dF_i(w) + \int_{-\infty}^{y(i)} \tilde{\Delta}r(i,w)dF_i(w), \tag{4.72}$$

$$P(j|i,\mathcal{D}(i)) = p(j|i,1) + \tilde{\Delta}p(j|i)F_i(y(i)), \tag{4.73}$$

and a *policy improvement routine* is to solve for a certain given $v(i)$, $i \in \mathcal{I}$, the maximization problem

$$g(i) = \max_{y(i)} \int_{-\infty}^{y(i)} \tilde{\Delta}V(i,w)dF_i(w). \tag{4.74}$$

Lemma 3 The solutions of the maximization problem of (4.71) and (4.74) are both given by $y(i) = c(i)$.

Proof See Appendix B.

■ Separable Case

In the case, from (4.30)we have

$$\tilde{\Delta}V_t(i,w) = \tilde{\Delta}r(i)w + \tilde{\Delta}z_t(i) = \tilde{\Delta}r(i)(w - c_t(i)) \tag{4.75}$$

where

$$c_t(i) = -\tilde{\Delta}z_t(i)/\tilde{\Delta}r(i), \tag{4.76}$$

which is the solution of $\tilde{\Delta}V_t(i,w) = 0$. Hence, from Corollary 3 we have

Case A If $r(i,x)$ is strictly increasing in x, then we have

$$v_t(i) = r(i,1)\mu_i + z_t(i,1) + \tilde{\Delta}r(i)T_i(c_t(i)) \tag{4.77}$$

Case B If $r(i,x)$ is strictly decreasing in x, then we have

$$v_t(i) = r(i,1)\mu_i + z_t(i,1) + \tilde{\Delta}r(i)S_i(c_t(i)). \tag{4.78}$$

Policy Iteration Algorithm VII

In the case we have

$$\tilde{\Delta}r(i,w) = \tilde{\Delta}r(i)w + \tilde{\Delta}e(i) = \tilde{\Delta}r(i)(w - h(i)) \tag{4.79}$$

where

$$h(i) = -\tilde{\Delta}e(i)/\tilde{\Delta}r(i). \tag{4.80}$$

Case A A *value determination operation* is to solve for a given $y(i)$, $i \in \mathcal{I}$, the system of equations (2.34) with (4.70) and

$$R(i, \mathcal{D}(i)) = r(i,1)\mu + e(i,1) + \tilde{\Delta}r(i)T_i(y(i), h(i)). \tag{4.81}$$

Case B A *value determination operation* is to solve for a given $y(i)$, $i \in \mathcal{I}$, the system of equations (2.34) with (4.73) and

$$R(i, \mathcal{D}(i)) = r(i,1)\mu + e(i,1) + \tilde{\Delta}r(i)S(y(i), h(i)). \tag{4.82}$$

In both cases, *policy improvement routines* are reduced to setting $y(i) = c(i)$ from Lemma 3.

5.3 An Enforced Restriction

All the discussions so far have been made on the assumption that $V_t(i, w, x)$ is either *concave* or *convex* in x. Here we eliminate the assumption, and let us restrict partitions $\mathcal{D}(i) = \{D(i,x) \mid x \in \mathcal{A}(i)\}$ of Ω_i on $\mathcal{A}(i)$, prescribing the optimal decision rule, only to ones that have the following property $\mathcal{P}(i)$:

$$D(i,x) = Z(i,x) - Z(i,x+1), \quad 1 \le x \le k(i), \tag{4.83}$$

where $Z(i,x)$ is nonincreasing in x with $Z(k(i),1) = \Omega_i$ and $Z(i, k(i)+1) = \phi_i$. In the next section, we will demonstrate two cases that such an enforced restriction has sufficient rationality. In the case, (2.22) can be rewritten as follows.

$$v_t(i) = \int_{w \in \Omega_i} V_t(i, w, 1) dF_i(w) + \max_{Z(i,2) \supseteq Z(i,3) \supseteq \cdots \supseteq Z(i,k(i))} \sum_{x=2}^{k(i)} \int_{w \in Z(i,x)} \Delta V_t(i, w, x) dF_i(w). \tag{4.84}$$

Then, the solution of the maximization problem in the right hand side is quite the same as (4.10) to (4.14).

• *A Further Restriction* As an example of further enforced restriction, we can consider the following case.

$$Z(i,x) = \{w \mid y(i,x) \le w\} \tag{4.85}$$

with $y(i,x) = (y(i,x,1), y(i,x,2), \cdots, y(i,x,M))'$ where $y(i,x)$ is nondecreasing in x for all i. Then, the above maximization problem can be reduced to

$$\max_{y(i,2) \le y(i,3) \le \cdots \le y(i,k(i))} \sum_{x=2}^{k(i)} \int_{y(i,x) \le w} \Delta V_t(i, w, x) dF_i(w). \tag{4.86}$$

In the case, letting

$$g_x(y(i,x)) = \int_{y(i,x) \le w} \Delta V_t(i, w, x) dF_i(w), \tag{4.87}$$

we have the following recurrent equation.

$$h_x(s) = \max_{y(i,x) \le s} \{g_x(y(i,x)) + h_{x-1}(y(i,x))\}, \quad 2 \le x \le k(i), \tag{4.88}$$

where $\boldsymbol{s} = (s_1, s_2, \cdots, s_M)'$. Hence, if $\boldsymbol{y}_t(\boldsymbol{i}, x, \boldsymbol{s})$, $x = 2, 3, \cdots, k(\boldsymbol{i})$, is the solution of (4.88), then the solution $\boldsymbol{y}_t^*(\boldsymbol{i}, x)$, $x = 2, 3, \cdots, k(\boldsymbol{i})$, of (4.86) becomes

$$\boldsymbol{y}_t^*(\boldsymbol{i}, x) = \begin{cases} \boldsymbol{y}_t(\boldsymbol{i}, k(\boldsymbol{i}), \infty), & x = k(\boldsymbol{i}), \\ \boldsymbol{y}_t(\boldsymbol{i}, x, \boldsymbol{y}^*(\boldsymbol{i}, x+1)), & x = 2, 3, \cdots, k(\boldsymbol{i}) - 1. \end{cases} \tag{4.89}$$

5. Applications

In this section, taking five examples, we demonstrate how for the models in the previous sections to be applied to concrete sequential stochastic decision problems.

5.1 Sequential Resource Allocation Problem

Consider there are available N units of resource, which are to be allocated among projects sequentially appearing one by one at each point in time over a given planning horizon where a project is assumed to be obtained with probability $p \leq 1$ at each point in time. Each obtained project have a value $w > 0$ where values $w, w', \cdots$ of successive projects are independently identically distributed random variables with a common known distribution function $F(w)$ having a sample space $\Omega = (0, \infty)$ and a finite expectation μ. Now, regarding no project being obtained as a fictitious project with value 0 being obtained, we can combine the p and $F(w)$ into a distribution function $H(w)$ whose probability density function is $h(w) = (1-p)I(w=0) + pf(w)I(w>0)$. Then, for any given function $s(w)$, we have

$$\int_0^\infty s(w)dH(w) = (1-p)s(0) + p\int_{0+}^\infty s(w)dF(w). \tag{5.1}$$

where the domains of integration $\int_0^\infty$ and $\int_{0+}^\infty$ are, respectively, $[0, \infty)$ and $(0, \infty)$.

When i units of resource remain, if x units of them are allocated to a project with value w where $0 \leq x \leq i$, then a return of $r(w, x)$ can be expected where i, N, and x are all nonnegative integers. Postulate that $r(w, x)$ is nondecreasing in w and x with $r(0, x) = r(w, 0) = 0$ and either concave or convex in x and that the difference $\Delta r(w, x)$ is either nondecreasing or nonincreasing in w. The objective here is to maximize the total expected present discounted return that will be gained over any given planning horizon where, if y units of resource remain at the end of the planning horizon, then it is evaluated as $\alpha(y)$.

Now it will be easily realized that a first-state of the decision process at each point in time is characterized by the amount i of resource then remaining and a second-state by a value w of a project then obtained, including a fictitious project. Hence, in the case, the space of first-states i is $\mathcal{I} = \{0, 1, \cdots, N\}$, an action space is $\mathcal{A}(i) = \{0, 1, \cdots, i\}$, an immediate reward is $r(i, w, x) = r(w, x)$, and the probability of a current first-state i changing into j at the next point in time is $p(j|i, x) = 1$ for $0 \leq j = i - x \leq i$ and $p(j|i, x) = 0$ for $i < j \leq N$.

Now, let $u_t(i, w)$ denote the maximum of the total expected present discounted return that will be gained up to time 0, starting from time t in state (i, w). Then, clearly

$$u_0(i, w) = \max_{0 \leq x \leq i}\{r(i, w) + \alpha(i - x)\}, \tag{5.2}$$

and (2.7) and (2.6) for $t \geq 1$ become, respectively,

$$u_t(i, w) = \max_{0 \leq x \leq i} V_t(i, w, x), \tag{5.3}$$

$$V_t(i,w,x) = r(w,x) + \beta \int_0^\infty u_{t-1}(i-x,\xi)dH(\xi). \tag{5.4}$$

Now define

$$v_t(i) = \int_0^\infty u_t(i,w)dH(w). \tag{5.5}$$

Then, (5.4) can be written

$$V_t(i,w,x) = r(w,x) + \beta v_{t-1}(i-x), \tag{5.6}$$

Lemma 4 *If $r(w,x)$ is concave (convex) in x, so also is $V_t(i,w,x)$ for all t.*

Proof See Appendix B.

- Concave Case

Suppose $r(w,x)$ is concave in x, hence so also is $V_t(i,w,x)$ for all t from Lemma 4. In the case, (4.3) becomes

$$\Delta V_t(i,w,x) = \Delta r(w,x) - \beta \Delta v_{t-1}(i-x+1) \tag{5.7}$$

for $0 \le x \le i$. Now, let a solution of $\Delta V_t(i,w,x) = 0$ with unknown w, if exists, be denoted by $c_t(i,x)$. Then, from Corollary 1 we have the following:

Case A Suppose $\Delta r(w,x)$ is nondecreasing in w. If the solution of $\Delta V_t(i,w,x) = 0$ does not exist, then let $c_t(i,x) = -\infty$. In the case, (4.15) becomes for $t \ge 1$

$$v_t(i) = \beta v_{t-1}(i) + \sum_{x=1}^{i} \int_{c_t(i,x)}^{\infty} (\Delta r(w,x) - \beta \Delta_{t-1}(i-x+1))dF(w). \tag{5.8}$$

Case B Suppose $\Delta r(w,x)$ is nonincreasing in w. If the solution of $\Delta V_t(i,w,x) = 0$ does not exist, then let $c_t(i,x) = \infty$. In the case, (4.16) becomes for $t \ge 1$

$$v_t(i) = \beta v_{t-1}(i) + \sum_{x=1}^{i} \int_{-\infty}^{c_t(i,x)} (\Delta r(w,x) - \beta \Delta_{t-1}(i-x+1))dF(w). \tag{5.9}$$

Now, the optimal decision rule can be described as follows. In ***Case A*** (***Case B***), if the value w of an project obtained at time t when i units of resource remain is in $(c_t(i,x), c_t(i,x+1)]$ ($[c_t(i,x+1), c_t(i,x))$), then invest x units of resource where $c_t(i,0) = -\infty$ and $c_t(i,i+1) = \infty$ ($c_t(i,0) = \infty$ and $c_t(i,i+1) = -\infty$).

Assume $r(w,x) = r(x)w$ where $r(x)$ is concave and strictly increasing in x with $r(0) = 0$. Then, from (4.33) and (4.36) we have, noticing $z_t(i,x) = -\beta \Delta v_{t-1}(i-x+1)$ in the case,

$$c_t(i,x) = \beta \Delta v_{t-1}(i-x+1)/\Delta r(x), \tag{5.10}$$

$$v_t(i) = \beta v_{t-1}(i) + \sum_{x=1}^{i} \Delta r(x) T(c_t(i,x)). \tag{5.11}$$

- Convex Case

Suppose $r(w,x)$ is convex in x, hence so also is $V_t(i,w,x)$ for all t from Lemma 4. In the case, (4.60) becomes

$$\tilde{\Delta} V_t(i,w) = \tilde{\Delta} r(w,i) - \beta \tilde{\Delta} v_{t-1}(i), \tag{5.12}$$

Now, let a solution of equation $\Delta V_t(i,w) = 0$ with unknown w, if exists, be denoted by $c_t(i)$. Then, from Corollary 3 we have the following:

Case A Suppose $\tilde{\Delta}r(w,x)$ is nondecreasing in w. If the solution of $\tilde{\Delta}V_t(i,w) = 0$ does not exist, then let $c_t(i) = -\infty$. In the case,

$$v_t(i) = \beta v_{t-1}(i) + \int_{c_t(i)}^{\infty} (\tilde{\Delta}r(w,i) - \beta\tilde{\Delta}v_{t-1}(i))dF(w). \tag{5.13}$$

Case B Suppose $\tilde{\Delta}r(w,x)$ is nonincreasing in w. If the solution of $\tilde{\Delta}V_t(i,w) = 0$ does not exist, then let $c_t(i) = \infty$. In the case,

$$v_t(i) = \beta v_{t-1}(i) + \int_{-\infty}^{c_t(i)} (\tilde{\Delta}r(w,i) - \beta\tilde{\Delta}v_{t-1}(i))dF(w). \tag{5.14}$$

Now, the optimal decision rule can be prescribed as follows. In ***Case A*** (***Case B***), if the value w of an project obtained at time t when i unit of resource remain is in $(-\infty, c_t(i)]$ ($[c_t(i), \infty)$), then invest no unit of resource, or else the whole remaining resource.

If $r(w,x) = r(x)w$ where $r(x)$ is convex and strictly increasing in x with $r(0) = 0$, then, from (4.76) and (4.77), we have, noticing $\tilde{z}_t(i) = -\beta\tilde{\Delta}v_{t-1}(i)$,

$$c_t(i) = \beta\tilde{\Delta}v_{t-1}(i)/\tilde{\Delta}r(i), \tag{5.15}$$

$$v_t(i) = \beta v_{t-1}(i) + \tilde{\Delta}r(i)T(c_t(i)) \tag{5.16}$$

where $\tilde{\Delta}r(i) = r(i) - r(0)$.

5.2 Inventory Problem I

Consider certain products which are manufactured from price fluctuating material. Here, let a unit of the material is used in the production of each product, and let purchasing price w per unit material fluctuates from day to day where subsequent purchasing prices $w, w', \cdots$ are independent identically distributed random variables having a known distribution function $F(w)$ with a finite expectation μ_F. Let a selling price p of a product be constant. Now, assume that every day any quantity of material can be purchased; however, the capacity of the warehouse in which the material is stocked is of N units, so the amount of material that can be purchased every day is limited by the capacity. Furthermore assume that a policy of selling as many products as possible for demand of each day is employed. Here postulate that required production time of the product is very short, so, for convenience, let it be zero, or instant. Demands for products on subsequent days, $\theta, \theta', \cdots$, are assumed to be independent identically distributed random variables having a known discrete distribution function $G(\theta)$, $\theta = 0, 1, \cdots$, with a finite expectation μ_G, so the probability of θ is $g(\theta) = G(\theta) - G(\theta - 1)$. Let K be a fixed ordering cost of the material, and let an inventory holding cost per unit per day for the material, cost incurred if one unit is held over to the next day, be denoted by h. Finally, assume that there exists no shortage cost and that, if a stockout of product occurred in the earlier days, then customers are not willing to wait, hence their demands are lost. The objective here is to find the optimal purchasing policy prescribing what quantity of material to purchase every day so as to maximize the total expected present discounted net profit over infinite planning horizon.

Let i represent the inventory of material at the beginning of each day: this is a first-state of the inventory system, so $\mathcal{I} = \{0, 1, \cdots, N\}$. Let q denote the purchasing quantity of the material of each day. Then, the range of the possible q on a day with inventory i is

$0 \le q \le N - i$. Now, let $x = i + q$, the level to which the current inventory i increases by purchasing q units, so the range of the possible x is $i \le x \le N$. Here, it goes without saying that we may take any one of q and x as a decision variable. For convenience, we shall take the latter x, so an action space becomes $\mathcal{A}(i) = \{i, i+1, \cdots, N\}$.

Now, if $x - i$ units are purchased on a day with inventory i, then an inventory in the morning of the next day, provided that the demand of the product on the day is of θ units, becomes $j = \max\{x - \theta, 0\}$. Here, letting $\mathcal{S}(j|i,x)$ be the set of θ for which the current inventory i changes into j on the next day, provided that $x - i$ units are purchased, we have

$$\mathcal{S}(j|i,x) = \{\theta \mid j = \max\{x - \theta\}\} \tag{5.17}$$

for $0 \le i, j \le N$ and $i \le x \le N$. Therefore, the probability that the current inventory i changes into j on the next day becomes $p(j|i,x) = \Pr\{\theta \in \mathcal{S}(j|i,x)\}$, that is,

$$p(j|i,x) = \begin{cases} g(x) + g(x+1) + \cdots & j = 0, \\ g(x-j) & 0 < j \le x, \\ 0 & x < j \le N. \end{cases} \tag{5.18}$$

Now, suppose $x - i$ units are purchased on a day with inventory i. Then, the expected net profit gained on the day can be expressed as

$$r(i,w,x) = \sum_{\theta=0}^{\infty} \Big(p\min\{x,\theta\} - w(x-i) - K\delta_{x-i} - h\max\{x-\theta,0\}\Big) g(\theta) \tag{5.19}$$

where, in general, $\delta_a = 1$ if $a > 0$, or else $\delta_a = 0$. Then, (5.19) can be arranged as

$$r(i,w,x) = (x-i)w + e(i,x) \tag{5.20}$$

where

$$e(i,x) = -K\delta_{x-i} + \sum_{\theta=0}^{\infty}(p\min\{x,\theta\} - h\max\{x - \theta\, 0\})g(\theta), \tag{5.21}$$

hence, for $i + 1 \le x \le N$ we have $\Delta r(i,x) = -1$ and

$$\Delta e(i,x) = p(1 - G(x-1)) + K(\delta_{x-i-1} - 1) - hG(x-1). \tag{5.22}$$

An Enforced Restriction of Partition Space *In the inventory system, it will not be so nonrational to think that the lower the purchasing price w may be, the more quantity should be purchased, and vice versa. This leads us to the purchasing rule with property $\mathcal{P}(i)$ such that, for a given nonincreasing sequence $y(i,x)$, $i = i, i+1, \cdots, N+1$, with $y(i,i) = \infty$ and $y(i, N+1) = -\infty$, if $w \in D(i,x) = [y(i,x+1), y(i,x))$, then purchase $x - i$ units. Then $\mathcal{D}(i) = \{D(i,x) \mid x = i, i+1, \cdots, N\}$ is a partition of $\Omega = (-\infty, \infty)$ on $\mathcal{A}(i)$, and let $\Lambda(i|\mathcal{P}(i))$ be a partition space consisting of all such partitions.*

Now, below, let us consider only purchasing rules such as prescribed above. Then, let $u(i,w)$ be the maximum of the total expected present discounted net profit gained over infinite planning horizon starting from a day with buying price w and inventory i. Then

$$V(i,w,x) = (x-i)w + z(i,x) \tag{5.23}$$

where

$$z(i,x) = e(i,x) + \beta \sum_{\theta=0}^{\infty} v(\min\{x-\theta, 0\})g(\theta), \tag{5.24}$$

$$v(i) = \int_0^{\infty} u(i,w)dF(w). \tag{5.25}$$

In the case,

$$\Delta z(i,x) = \Delta e(i,x) + \beta \sum_{\theta=0}^{x-1} \Delta v(x-\theta)g(\theta). \tag{5.26}$$

Then, (4.42) and (4.24) becomes, respectively,

$$R(i,\mathcal{D}(i)) = \sum_{\theta=0}^{\infty}(p\min\{i,\theta\} - h\max\{i-\theta,0\})g(\theta) - \sum_{x=i+1}^{N} S(y(i,x), \Delta e(i,x)), \tag{5.27}$$

$$P(j|i,\mathcal{D}(i)) = p(j|i,i) + \sum_{x=i+1}^{N} \Delta p(j|i,x)F(y(i,x)). \tag{5.28}$$

Consequently, a *value determination operation* is to solve for a given $\mathcal{D}(i) \in \Lambda(i|\mathcal{P}(i))$, $0 \le i \le N$, the system of equations (2.34) with (5.27) and (5.28), and a *policy improvement routine* is to solve for a given $v(i)$, $0 \le i \le N$, the recurrent equation (4.25), i.e., in the case

$$h_x(s) = \max_{y(i,x)\ge s}\{g_x(y(i,x)) + h_{x-1}(y(i,x))\}, \quad i+1 \le x \le N, \tag{5.29}$$

with $h_i(s) \equiv 0$ where $g_x(y(i,x)) = -S(y(i,x), \Delta z(i,x))$ from (4.43) and (4.33).

5.3 Inventory Problem II

In the inventory problem of the previous section, suppose selling price p of the product is also varies from day to day. Below let us denote purchasing price and selling price by, respectively, w_1 and w_2, and let the vector $\boldsymbol{w} = (w_1, w_2)$, called *price vector*. Assume that subsequent price vectors $\boldsymbol{w}, \boldsymbol{w}', \cdots$ are independent identically distributed random variable having a distribution function $F(\boldsymbol{w})$. In this case, the expected net profit gained on each day, provided that x units of material are purchased, can be expressed by

$$r(i,\boldsymbol{w},x) = \sum_{\theta=0}^{\infty}\{w_2\min\{x,\theta\} - w_1(x-i) - K\delta_{x-i} - h\max\{x-\theta,0\}\}g(\theta), \tag{5.30}$$

from which we have

$$\Delta r(i,w,x) = -w_1 + w_2(1-G(x-1)) + K(\delta_{x-i-1} - 1) - hG(x-1). \tag{5.31}$$

An Enforced Restriction of Partition Space *Also in this problem, we are led to the following purchasing rule for any given* w_2. *For a given nonincreasing sequence* $y(i,x,w_2)$, $x = i, i+1, \cdots, N+1$, *with* $y(i,i,w_2) = \infty$ *and* $y(i,N+1,w_2) = -\infty$, *if* $w_1 \in D(i,x) = \{w_1 \mid y(i,x+1,w_2) < w_1 \le y(i,x,w_2)\}$, *then purchase* $x-i$ *units. Then,* $\mathcal{D}(i) = \{D(i,x) \mid x \in \mathcal{A}(i)\}$ *is a partition of* $\Omega = (-\infty,\infty)$ *on* $\mathcal{A}(i)$, *and let* $\Lambda(i|\mathcal{P}(i))$ *be a partition space consisting of all such partitions. Furthermore, we, if want, may impose the following restriction:* $y(i,x,w_2) = a(i,x)w_2 + b(i,x)$ *with* $a(i,i) = b(i,i) = \infty$ *and* $a(i,N+1) = b(i,N+1) = -\infty$ *where both* $a(i,x)$ *and* $b(i,x)$ *are nonincreasing in* x *and nonnegative for all* i *and* x.

Now, let us consider only purchasing rules such as prescribed above. Then, let $u(i, \boldsymbol{w})$ be the maximum of the total expected present discounted net profit over infinite planning horizon, starting from a day with price vector $\boldsymbol{w}$ and inventory i, and let

$$v(i) = \int_0^\infty u(i, \boldsymbol{w}) dF(\boldsymbol{w}). \tag{5.32}$$

Then (2.25) and (2.26) becomes, respectively,

$$R(i, \mathcal{D}(i)) = \sum_{x=i}^{N} \int_0^\infty \int_{y(i,x+1,w_2)}^{y(i,x,w_2)} r(i, \boldsymbol{w}, x) dF(\boldsymbol{w}) \tag{5.33}$$

$$P(j|i, \mathcal{D}(i)) = \sum_{x=i}^{N} p(j|i, x) \int_0^\infty \int_{y(i,x+1,w_2)}^{y(i,x,w_2)} dF(\boldsymbol{w}) \tag{5.34}$$

where $p(j|i, x)$ is the same as (5.18) and the first and second integrations are with respect to w_2 and w_1, respectively. Then, the above two expressions can be rewritten as follows.

$$R(i, \mathcal{D}(i)) = \int_0^\infty \int_0^\infty r(i, \boldsymbol{w}, i) dF(\boldsymbol{w}) + \sum_{x=i+1}^{N} \int_0^\infty \int_0^{y(i,x,w_2)} \Delta r(i, \boldsymbol{w}, x) dF(\boldsymbol{w}) \tag{5.35}$$

$$= \sum_{\theta=0}^{\infty} \{\mu_2 \min\{i, \theta\} - h \max\{i - \theta, 0\}\} g(\theta)$$

$$+ \sum_{x=i+1}^{N} \int_0^\infty \int_0^{y(i,x,w_2)} \Delta r(i, \boldsymbol{w}, x) dF(\boldsymbol{w}) \tag{5.36}$$

$$P(j|i, \mathcal{D}(i)) = p(j|i, i) + \sum_{x=i+1}^{N} \Delta p(j|i, x) \int_0^\infty \int_0^{y(i,x,w_2)} dF(\boldsymbol{w}), \tag{5.37}$$

where μ_2 is an expectation of w_2.

Then a *value determination operation* is to solve the system of equations (2.34) with (5.36) and (5.37) for a certain given nondecreasing sequence $y(i, x, w_2)$, $x = i, i+1, \cdots, N$, and a *policy improvement routine* is reduced to the following maximization problem.

$$\max_{y(i,i,\cdot) \geq y(i,i+1,\cdot) \geq \cdots \geq y(i,N+1,\cdot)} \{R(i, \mathcal{D}(i)) + \beta \sum_{j=1}^{N} P(j|i, \mathcal{D}(i)) v(j)\}, \tag{5.38}$$

which, by substituting (5.35) and (5.37) into , can be reduced to

$$\max_{y(i,i+1,\cdot) \geq y(i,i+1,\cdot) \geq \cdots \geq y(i,N,\cdot)} \sum_{x=i+1}^{N} g_x(y(i, x, \cdot)) \tag{5.39}$$

where, in general,

$$g_x(z(\cdot)) = \int_0^\infty \int_0^{z(w_2)} \left(\Delta r(i, \boldsymbol{w}, x) + \beta \sum_{j=0}^{N} \Delta p(j|i, x) v(j)\right) dF(\boldsymbol{w}). \tag{5.40}$$

If $y(i, x, w_2) = a(i, x) w_2 + b(i, x)$, then (5.39) becomes

$$\max_{\substack{a(i,i) \geq a(i,i+1) \geq \cdots \geq a(i,N+1) \\ b(i,i) \geq b(i,i+1) \geq \cdots \geq b(i,N+1)}} \sum_{x=i+1}^{N} g_x(a(i,x), b(i,x)) \tag{5.41}$$

where

$$g_x(a,b) = \int_0^\infty \int_0^{aw_2+b} \Big(\Delta r(i,w,x) + \beta \sum_{j=0}^{N} \Delta p(j|i,x) v(j)\Big) dF(w). \tag{5.42}$$

The above maximization problem can be solved as a standard dynamic programming problem.

5.4 Customers Selection Problem

Consider the following discrete-time queuing process with a single service station in which more than N customers can not be accommodated and service for a customer at a certain point in time is completed in the next point in time with a given probability p. Given that l customers, $l = 0, 1, \ldots$, arrived at a certain point in time, let q_{lm}, $m = 0, 1, \cdots$, be the probability of m customers arriving at the next point in time. Each of l customers arriving at a certain point in time has a value $w_{ln} \in (0, \infty)$, $n = 1, 2, \cdots, l$, assumed to be mutually independent. Here let $w_{l0} = 0$ for all l, which should be regarded as the value of a dummy customer. Now let $\boldsymbol{w}_l = (w_{l0}, w_{l1}, \cdots .w_{ll})'$, and assume that subsequent $\boldsymbol{w}_l, \boldsymbol{w}'_{l'}, \cdots$ are also stochastically independent. That is, values of all customers arriving at not only each point in time but also subsequent point in times are assumed to be mutually independent, and furthermore assume that values of these customers are random variable having a known identical distribution function $F(w)$ with a finite expectation μ. If some of customers arriving at each point in time are accepted, then the sum of their values can be obtained. The objective here is to find the optimal decision rule prescribing how many of arriving customers to accept at each point in time so as to maximize the total expected present discounted value obtained over any given finite planning horizon.

Now, a state of the process at any point in time can be characterized by the following three entities: the number i of customers in the system, the number l of customers that have just arrived, and the vector $\boldsymbol{w}_l = (w_{l1}, w_{l2}, \cdots, w_{ll})$ of values of the customers. In this case, for convenience, we may regard the pair (i, l) as a first-state and the $\boldsymbol{w}_l$ as a second-state. Now, the range of the number of customers that can be accepted when in first state (i, l) is

$$0 \leq x \leq k(i,l) \stackrel{d}{=} \min\{N - i, l\}. \tag{5.43}$$

Here let us rearrange the values $w_{l1}, w_{l2}, \cdots, w_{ll}$ of arriving customers as $w_{l1} \geq w_{l2} \geq \cdots \geq w_{ll}$, and newly define the vector $\boldsymbol{w}_l = (w_{l1}, w_{l2}, \cdots, w_{ll})$. Now, in the case, it goes without saying that, if x ones of l customers arriving at a certain point in time are accepted, then the best x ones in an order of their value size should be accepted, so, if x customers are accepted at a time in first-state (i, l), then the total value obtained at that time becomes

$$r(i, l, \boldsymbol{w}_l, x) = \sum_{n=0}^{x} w_{ln}, \tag{5.44}$$

and the probability of the i changing into j at the next point in time is given by, for $0 \leq x \leq k(i, x)$ and $0 \leq i \leq N$,

$$p(j|i,x) = \begin{cases} \begin{cases} 1 & j=0, \\ 0 & j=\text{else}, \end{cases} & |i|+|x|=0, \\ \begin{cases} 1-p & j=i+x, \\ p & j=i+x-1, \\ 0 & j=\text{else}, \end{cases} & |i|+|x|\neq 0. \end{cases} \tag{5.45}$$

Now, let $u_t(i,l,\boldsymbol{w}_l)$ represent the maximum of the total expected present discounted value obtained from customers that will be accepted over the given planning horizon, starting from time $t \geq 0$ in state $(i,l,\boldsymbol{w}_l)$,where let $u_{-1}(i,l,\boldsymbol{w}_l) \equiv 0$. Here define

$$v_t(i,l) = \int_0^\infty \int_0^\infty \cdots \int_0^\infty u_t(i,l,\boldsymbol{w}_l) dF(w_{l1}) dF(w_{l2}) \cdots dF(w_{ll}). \tag{5.46}$$

Then

$$u_t(i,l,\boldsymbol{w}_l) = \max_{0 \leq x \leq k(i,l)} V_t(i,l,\boldsymbol{w}_l,x) \tag{5.47}$$

where from (4.46) and (4.47) we have

$$V_t(i,l,\boldsymbol{w}_l,x) = r(i,l,\boldsymbol{w}_l,x) + z_t(i,l,x), \tag{5.48}$$

$$z_t(i,l,x) = \beta \sum_{m=0}^{\infty} q_{lm}\Big((1-p)v_{t-1}(i+x,m) + pv_{t-1}(\max\{i+x-1,0\},m)\Big). \tag{5.49}$$

Hence, (4.49) becomes, for $1 \leq x \leq k(i,l)$,

$$c_t(i,l,x) = -\beta \sum_{m=0}^{\infty} q_{lm}((1-p)\Delta v_{t-1}(i+x,m) + A) \tag{5.50}$$

where

$$A = \begin{cases} 0 & \text{if } |i|+|x|=0 \\ p\Delta v_{t-1}(i+x-1,m) & \text{if } |i|+|x| \neq 0 \end{cases} \tag{5.51}$$

Lemma 5 *For all $t \geq 0$, $V_t(i,l,\boldsymbol{w}_l,x)$ is concave in x for all i, l, and $\boldsymbol{w}_l$, and $v_t(i,l)$ is nonincreasing and concave in i.*

Proof See Appendix B.

From the above lemma and (4.52), the optimal partition $C_t(i,l)$ of $\Omega_l = \prod_{n=1}^{l}(0,\infty)$ on $\mathcal{A}(i,l) = \{0,1,\cdots,k(i,l)\}$ is given by, for $0 \leq x \leq k(i,l)$,

$$C_t(i,l,x) = \{\boldsymbol{w}_l \mid c_t(i,l,x) < w_{lx}, w_{l,x+1} \leq c(i,l,x+1)\} \tag{5.52}$$

where $c(i,l,x)$ is nonnegative and nondecreasing in i and x from Lemma 5 with $c_t(i,l,0) = -\infty$, $c_t(i,l,k(i,l)+1) = \infty$, and $w_{l,k(i,l)+1} = -\infty$. Thus, the optimal customers selection rule can be stated as follows: if $c_t(i,l,x) < w_{lx}$ and $w_{l,x+1} \leq c_t(i,l,x+1)$, then accept the best x customers. Furthermore we have from (4.53)

$$v_t(i,l) = \beta \sum_{m=0}^{\infty} q_{lm}\Big((1-p)v_{t-1}(i,m) + pv_{t-1}(\max\{i-1,0\},m)\Big) + \sum_{x=1}^{k(i,l)} T_{lx}(c_t(i,l,x)) \tag{5.53}$$

Finally, the policy iteration algorithm is quite the same as (4.54) to (4.58) except that the vector $\boldsymbol{i}$ is replaced with a scalar i.

Appendix A

For expressional simplicity, in the proofs below, including Appendixes B and C, we shall drop the symbols "t", "i", and "$w \in \Omega$" in $\int_{w\in\Omega}$ from all the related symbols.

Proof of Theorems 1 First, note that (4.1) can be expressed as

$$C(x) = \{w \mid \textstyle\sum_{n=a+1}^{x} \Delta V(w,n) > 0,\ 1 \le a < x \\ \textstyle\sum_{n=x+1}^{b} \Delta V(w,n) \le 0,\ x < b \le k\}, \quad 1 \le x \le k.$$

Define $\widetilde{C}(x) = \{w|\Delta V(w,x) > 0 \ge \Delta V(w,x+1)\}$ where $\Delta V(w,1) = \infty$ and $\Delta V(w,k+1) = -\infty$. Then clearly $C(x) \subset \widetilde{C}(x)$. By concavity assumption, if $\Delta V(w,x) > 0$, then $\Delta V(w,n) > 0$ for all $n \le x$; hence, $\sum_{n=a+1}^{x} \Delta V(w,n) > 0$ for all a such as $1 \le a < x$, and if $\Delta V(w,x+1) \le 0$, then $\Delta V(w,n) \le 0$ for all $n \ge x+1$, hence $\sum_{n=x+1}^{b} \Delta V(w,n) \le 0$ for all b such as $x < b \le k$. Thus $\widetilde{C}(x) \subset C(x)$, hence it follows that $\widetilde{C}(x) = C(x)$, i.e., (4.4) holds true. Now, since $C(x)$ $= \{w \mid \Delta V(w,x) > 0\} - \{w \mid \Delta V(w,x+1) > 0\}$, we have $I(w \in C(x)) = I(\Delta V(w,x) > 0) - I(\Delta V(w,x+1) > 0)$. Arranging (2.29) by substituting this into leads to

$$\begin{aligned} v &= \textstyle\sum_{x=1}^{k} \int (V(w,x)I(\Delta V(w,x) > 0) - V(w,x)I(w \in \Delta V(w,x+1) > 0))dF(w) \\ &= \textstyle\sum_{x=1}^{k} \int V(w,x)I(\Delta V(w,x) > 0)dF(w) \\ &\qquad - \textstyle\sum_{x=2}^{k+1} \int V(w,x-1)I(\Delta V(w,x) > 0))dF(w) \\ &= \textstyle\int V(w,1)I(\Delta V(w,1) > 0)dF(w) - \int V(w,k)I(\Delta V(w,k+1) > 0)dF(w) \\ &\qquad + \textstyle\sum_{x=2}^{k} \int V(w,x)I(\Delta V(w,x) > 0)dF(w). \end{aligned}$$

Noting $I(\Delta V(w,1) > 0) = 1$ and $I(\Delta V(w,k+1) > 0) = 0$, the above expression can be arranged as follows.

$$\begin{aligned} v &= \textstyle\int V(w,1)dF(w) + \sum_{x=2}^{k} \int V(w,x)I(\Delta V(w,x) > 0)dF(w) \\ &= \textstyle\int_{w\in Y(x)} V(w,1)dF(w) + \sum_{x=2}^{k} \int_{w\in Y(x)} V(w,x)dF(w). \end{aligned}$$

Proof of Theorem 2 Let $\widetilde{C}(1) = \{w|\Delta V(w) \le 0\}$ and $\widetilde{C}(k) = \{w|\Delta V(w) > 0\}$, which can be expressed as, respectively,

$$\widetilde{C}(1) = \{w \mid \textstyle\sum_{n=2}^{k} \Delta V(w,n) \le 0\}, \qquad \widetilde{C}(k) = \{w \mid \sum_{n=2}^{k} \Delta V(w,n) > 0\}.$$

Now, since $C(1)$ and $C(k)$ can be rewritten, respectively,

$$\begin{aligned} C(1) &= \{w \mid \textstyle\sum_{n=2}^{b} \Delta V(w,n) \le 0,\ 1 < b \le k\}, \\ C(k) &= \{w \mid \textstyle\sum_{n=a+1}^{k} \Delta V(w,n) > 0,\ 1 \le a < k\}, \end{aligned}$$

it follows that $C(1) \subset \widetilde{C}(1)$ and $C(k) \subset \widetilde{C}(k)$.

Now assume $\sum_{n=2}^{k} \Delta V(w,n) \leq 0$. In the case, if $\sum_{n=2}^{k-1} \Delta V(w,n) > 0$, then $\Delta V(w,n) > 0$ for at least one n such as $2 \leq n \leq k-1$, hence $\Delta V(w,k) \geq \Delta V(w,k-1) > 0$ because $\Delta V(w,n)$ is nondecreasing in n. This leads to the contradiction of $\sum_{n=2}^{k} \Delta V(w,n) > 0$. Accordingly it must be that $\sum_{n=2}^{k-1} \Delta V(w,n) \leq 0$. Similarly we can show in general $\sum_{n=2}^{b} \Delta V(w,n) \leq 0$ for all b such as $1 < b \leq k$, which implies $\tilde{C}(1) \subset C(1)$. Next, assume $\sum_{n=2}^{k} \Delta V(w,n) > 0$. In the case, if $\sum_{n=3}^{k} \Delta V(w,n) \leq 0$, then $\Delta V(w,n) \leq 0$ for at least one n such as $3 \leq n \leq k$, hence $\Delta V(w,2) \geq \Delta V(w,3) > 0$, which yields the contradiction of $\sum_{n=2}^{k} \Delta V(w,n) \leq 0$. Therefore it follows that $\sum_{n=3}^{k} \Delta V(w,n) \leq 0$. Similarly we can show in general $\sum_{n=a+1}^{k} \Delta V(w,n) \leq 0$ for all b such as $1 \leq a \leq k$, which implies $\tilde{C}(k) \subset C(k)$. Thus, it follows that $\tilde{C}(1) = C(1)$ and $\tilde{C}(k) = C(k)$. Furthermore, since $C(1) \bigcup C(k) = \Omega$, it eventually follows that $C(x) = \phi$ for $1 < x < k$. Consequently, noticing $I(w \in C(1)) = 1 - I(w \in C(k))$, we have from (2.29)

$$\begin{aligned} v &= \int V(w,1) I(w \in C(1)) dF(w) + \int V(w,k) I(w \in C(k)) dF(w) \\ &= \int \Delta V(w,1) dF(w) + \int_{w \in \Omega} V(w) I(w \in C(k)) dF(w) \\ &= \int V(w,1) dF(w) + \int_{w \in Y} V(w) dF(w). \end{aligned}$$

Appendix B

Proof of Lemma 1 *Proof of C being a partition of Ω on $\mathcal{A}$.* First, for any $n > m$ we have $C(x_n) \subset \{w | V(w,x_n) > V(w,x_m)\}$ and $C(x_m) \subset \{w | V(w,x_m) \geq V(w,x_n)\}$, hence, $C(x_n)$ and $C(x_m)$ are mutually exclusive for any $n > m$ or $m > n$. Next, we shall prove $\bigcup_{n \in \mathcal{A}^*} C(x_n) = \Omega$. For this, it suffices to show that any $w \in \Omega$ is contained in any one of $C(x_1), C(x_2), \cdots$, and $C(x_k)$. For any given $w \in \Omega$ let

$$V(w,x_{d_1}) = V(w,x_{d_2}) = \cdots = V(w,x_{d_N}) > \max\{V(w,x_{d_{N+1}}), V(w,x_{d_{N+2}}), \cdots, V(w,x_{d_k})\}$$

where $\{d_1, d_2, \cdots, d_N, d_{N+1}, d_{N+2}, \cdots, d_k\}$ is a permutation of $1, 2, \cdots, k$ with $d_1 < d_2 < \cdots < d_N$ and $d_{N+1} < d_{N+2} < \cdots < d_k$. Then, the following can be said:

1. If $d_1 = 1$, then $V(w,x_1) \geq V(w,x_j)$ for any $1 < j \leq k$, so $w \in C(x_1)$,
2. If $0 < d_1 < k$, then $V(w,x_{d_1}) > V(w,x_i)$ for any $1 \leq i < d_1$ and $V(w,x_{d_1}) \geq V(w,x_j)$ for any $d_1 < j \leq k$, so $w \in C(x_{d_1})$,
3. If $d_1 = k$, then $V(w,x_k) > V(w,x_i)$ for any $1 \leq i < k$, so $w \in C(x_k)$.

Thus it follows that the w is contained in one of $C(x_1), C(x_2), \cdots, C(x_k)$.

Proof of the C being an optimal partition of Ω on $\mathcal{A}$ and (2.13) holding. First, for any $i \in \Omega$, since $V(w,x_n) \leq u(w)$ for all n, from (2.12) with the C instead of $\mathcal{D}$ we get $V(w|C) \leq v(w) \sum_n I(w \in C(x_n)) = u(w) \cdots (1^*)$. Next, For any given $w \in \Omega$, if $w \in C(x_m)$ for a certain m, then $V(w|C) = V(w,x_m) \geq V(w,x_n)$ for all n; hence, it follows that $V(w|C) \geq u(w) \cdots (2^*)$. Thus, from (1^*) and (2^*) we have $v(w) = V(w|C)$.

Now, noticing that, in general, $I(A \bigcup B) = I(w \in A) + I(w \in B)$ for any two exclusive subsets A and B, for the partition C and any partition $\mathcal{D}$ we have

$$\begin{aligned} C(x_n) &= C(x_n) \bigcap \Omega = C(x_n) \bigcap (\bigcup_m D(x_m)) = \bigcup_m (C(x_n) \bigcap D(x_m)) \\ D(x_m) &= D(x_m) \bigcap \Omega = D(x_m) \bigcap (\bigcup_n C(x_n)) = \bigcup_n (C(x_n) \bigcap D(x_m)). \end{aligned}$$

Hence it follows that for any partition $\mathcal{D}$

$V(w|C) - V(w|\mathcal{D})$

$$= \sum_n V(w,x_n)I(w \in C(x_n)) - \sum_m V(w,x_m)I(w \in D(x_m))$$
$$= \sum_n \sum_m V(w,x_n)I(w \in C(x_n) \bigcap D(x_m)) - \sum_m \sum_n V(w,x_m)I(w \in C(x_n) \bigcap D(x_m))$$
$$= \sum_n \sum_m (V(w,x_n) - V(w,x_m))I(w \in C(x_n) \bigcap D(x_m)).$$

Since $V(w,x_n) - V(w,x_m) \geq 0$ for $w \in C(x_n) \cap D(x_m)$ $(\subset C(x_n))$, it follows that $V(w|\mathcal{C}) - V(w|\mathcal{D}) \geq 0$. Thus $V(w|\mathcal{C}) = \max_{\mathcal{D}\in\mathcal{A}} V(w|\mathcal{D})$; hence, $\mathcal{C}$ is the optimal partition of Ω on $\mathcal{A}$.

Proof of Lemma 2 Let λ = r.h.s of (2.20), i.e.,

$$\lambda = \max_{\mathcal{D}\in\mathcal{A}} \int V(w|\mathcal{D})dF(w).$$

Since $V(w|\mathcal{D}) \leq u(w)$ for any w and $\mathcal{D}$ from (2.16), we obtain

$$\lambda \leq \max_{\mathcal{D}\in\mathcal{A}} \int u(w)dF(w) = \int u(w)dF(w) = v.$$

On the other hand, clearly

$$\lambda \geq \int V(w|\mathcal{C})dF(w) = \int u(w)dF(w) = v.$$

Thus, it must be that $v = \lambda$.

Proof of Lemma 3 If $\widetilde{\Delta}V(w)$ is nondecreasing in w, then we have

$$\int_c^\infty \widetilde{\Delta}V(w)dF(w) - \int_y^\infty \widetilde{\Delta}V(w)dF(w) = \int_c^y \widetilde{\Delta}V(w)dF(w) \geq 0$$

in any case of $y \geq c$ or $c \geq y$. Accordingly $\int_y^\infty \widetilde{\Delta}V(w)dF(w)$ takes a maximum value at $y = c$. Almost the same as this for case that $\widetilde{\Delta}V(w)$ is nonincreasing in w.

Proof of Lemma 4 *Case that $r(w,x)$ is concave in x.* Then clearly $v_0(i)$ is also concave in i. Assume that $v_{t-1}(i)$ is concave in i. Then $V_t(i,w,x)$ is also concave in x. Now, for any integers $i,j \in [0,N]$ and any real number $\lambda \in [0,1]$ such that $\lambda i + (1-\lambda)j$ is an integer on $[0,N]$, we have

$$\begin{aligned}
u_t(\lambda i + \nu j, w) &= \max_{0\leq x\leq \lambda i+\nu j}\{r(x,w) + \beta v_{t-1}(\lambda i + \nu j - x)\} \quad \nu = 1-\lambda \\
&= \max_{0\leq \tilde{x}\leq \lambda i+\nu j}\{\tilde{r}(\tilde{x},w) + \beta \tilde{v}_{t-1}(\lambda i + \nu j - \tilde{x})\} \\
&= \max_{0\leq \lambda\tilde{y}+\nu\tilde{z}\leq \lambda i+\nu j}\{\tilde{r}(\lambda\tilde{y} + \nu\tilde{z},w) + \beta \tilde{v}_{t-1}(\lambda i + \nu j - \lambda\tilde{y} + \lambda\tilde{z}))\} \quad 0 \leq \tilde{y} \leq i, \ 0 \leq \tilde{y} \leq i \\
&\geq \max_{0\leq \lambda\tilde{y}+\nu\tilde{z}\leq \lambda i+\nu j}\{\lambda\tilde{r}(\lambda\tilde{y},w) + \nu\tilde{r}(\lambda\tilde{z},w) + \lambda\beta\tilde{v}_{t-1}(i-\tilde{y}) + \nu\beta\tilde{v}_{t-1}(i-\tilde{z}))\} \\
&\geq \max_{0\leq\tilde{y}\leq i, \ 0\leq\tilde{z}\leq j}\{\lambda\tilde{r}(\tilde{y},w) + \nu\tilde{r}(\tilde{z},w) + \lambda\beta\tilde{v}_{t-1}(i-\tilde{y}) + \nu\beta\tilde{v}_{t-1}(j-\tilde{z})\} \\
&= \lambda \max_{0\leq\tilde{y}\leq i}\{\tilde{r}(\tilde{y},w) + \beta\tilde{v}_{t-1}(i-\tilde{y})\} + \nu \max_{0\leq\tilde{z}\leq j}\{\tilde{r}(\tilde{z},w) + \beta\tilde{v}_{t-1}(i-\tilde{z}))\} \\
&= \lambda \max_{0\leq y\leq i}\{r(y,w) + v_{t-1}(i-y)\} + \nu \max_{0\leq z\leq j}\{r(z,w) + v_{t-1}(i-z))\} \\
&= \lambda u_t(i,w) + \nu u_t(j,w),
\end{aligned}$$

Thus, $u_t(i,w)$, hence $v_t(i)$ is concave in i.

Case that $r(w,x)$ is convex in x. Then clearly $v_0(i)$ is also convex in i. Suppose $v_{t-1}(i)$ is convex in i, hence $V_y(i,w,x)$. Consequently

$$u_t(i,w) = \max\{V_t(i,w,0), V_t(i,w,i)\} = \max\{r(w,0) + \beta v_{t-1}(i), r(w,i) + \beta v_{t-1}(0)\}$$

Now, for any integers $i, j \in [0, N]$ and any real number $\lambda \in [0,1]$ such that $\lambda i + (1-\lambda)j$ is an integer on $[0, N]$, we have

$$\begin{aligned}
u_t(\lambda i + \nu j, w) &= \max\{r(w,0) + \beta v_{t-1}(\lambda i + \nu j), r(w, \lambda i + \nu j) + \beta v_{t-1}(0)\} \\
&\leq \max\{r(w,0) + \lambda\beta v_{t-1}(i) + \nu\beta v_{t-1}(i), \lambda r(w,i) + \nu r(w,j) + \beta v_{t-1}(0)\} \\
&= \max\{\lambda(r(w,0) + \beta v_{t-1}(i)) + \nu(r(w,0) + \beta v_{t-1}(j)) \\
&\qquad\qquad \lambda(r(w,i) + \beta v_{t-1}(0)) + \nu(r(w,j) + \beta v_{t-1}(0))\} \\
&\leq \max\{\lambda(r(w,0) + \beta v_{t-1}(i)), \lambda(r(w,i) + \beta v_{t-1}(0))\} \\
&\qquad\qquad + \max\{\nu(r(w,0) + \beta v_{t-1}(j)), \nu(r(w,j) + \beta v_{t-1}(0))\} \\
&= \lambda \max\{r(w,0) + \beta v_{t-1}(i), r(w,i) + \beta v_{t-1}(0)\} \\
&\qquad\qquad + \nu \max\{r(w,0) + \beta v_{t-1}(j), r(w,j) + \beta v_{t-1}(0)\} \\
&= \lambda u_t(i,w) + \nu u_t(j,w)
\end{aligned}$$

Therefore $u_t(i,w)$, hence $v_t(i)$ is convex in i.

Proof of Lemma 5 Clearly $v_{-1}(i,l)$ is nonincreasing and concave in i for all n. Assume that $v_{t-1}(i,l)$ is nonincreasing and concave in i. Then $V_t(i,l,w^l,x)$ is nonincreasing in i. In addition to this, since $\min\{l, N-i\}$ is also nonincreasing in i, it follows from (5.47) that $u_t(i,l,w_l)$ is also nonincreasing in i, hence so also is $v_t(i,l)$. Now from (5.48) we have

$$\Delta V_t(0,l,w_l,1) = w_{l1} + \beta \textstyle\sum_{k=0}^{\infty} p_k(1-q)\Delta v_{t-1}(1,k) \qquad (1)$$

$$\Delta V_t(i,l,w_l,x) = w_{lx} + \beta \textstyle\sum_{k=0}^{\infty} p_k\Big((1-q)\Delta v_{t-1}(i+x,k) + q\Delta v_{t-1}(i+x-1,k)\Big) \qquad (2)$$

for all i, x such as $i^2 + (x-1)^2 \neq 0$. Then

$$\Delta V_t(0,l,w_l,2) = w_{l2} + \beta \textstyle\sum_{k=0}^{\infty} p_k\Big((1-q)\Delta v_{t-1}(2,k) + q\Delta v_{t-1}(1,k)\Big) \qquad (3)$$

Clearly $\Delta V_t(0,l,w_l,2) \leq \Delta V_t(0,l,w_l,1)$ and $\Delta V_t(0,l,w_l,x+1) \leq \Delta V_t(0,l,w_l,x)$ for all i, x such as $i^2 + (x-1)^2 \neq 0$. Thus, it follows that $V_t(i,l,w_l,x)$ is concave in x for all i, l, and w_l. Now we shall write $u_t(i,l,w_l)$ as follow.

$$\begin{aligned}
&u_t(i,l,w_l) = g(x) + H_t(i+x) \\
&g(x) = \textstyle\sum_{n=0}^{x} w_{ln} \\
&H_t(i+x) = \beta \textstyle\sum_{k=0}^{\infty} p_k\Big((1-q)v_{t-1}(i+x,k) + qv_{t-1}(\max\{i+x-1,0\},k)\Big)
\end{aligned}$$

Then (5.47) can be expressed as

$$u_t(i,l,w_l) = \max_{0 \leq x \leq \min\{l, N-i\}} \{g(x) + H_t(i+x)\}$$

where $g(x)$ and $H_t(i+x)$ are both concave function. Now we have

$$\begin{aligned}
u_t(\lambda i + \nu j, l, w_l) &= \max_{0 \leq x \leq \min\{l, N-\lambda i - \nu j\}} \{g(x) + H_t(\lambda i + \nu j + x)\} \\
&= \max_{0 \leq \tilde{x} \leq \min\{l, N-\lambda i - \nu j\}} \{\tilde{g}(\tilde{x}) + \widetilde{H}_t(\lambda i + \nu j + \tilde{x})\} \\
&= \max_{0 \leq \lambda\tilde{y} + \nu\tilde{z} \leq \min\{l, \lambda(N-i) + \nu(N-j)\}} \{\tilde{g}(\lambda\tilde{y} + \nu\tilde{z}) + \widetilde{H}_t(\lambda i + \nu i + \lambda\tilde{y} + \nu\tilde{z})\}
\end{aligned}$$

$$
\begin{aligned}
&\geq \max_{0\leq\lambda\tilde{y}+\nu\tilde{z}\leq\lambda\min\{l,N-i\}+\nu\min\{l,N-j\}}\{\tilde{g}(\lambda\tilde{y}+\nu\tilde{z})+\widetilde{H}_t(\lambda(i+\tilde{y})+\nu(i+\tilde{z}))\}\\
&\geq \max_{0\leq\lambda\tilde{y}+\nu\tilde{z}\leq\lambda\min\{l,N-i\}+\nu\min\{l,N-j\}}\{\lambda(\tilde{g}(\tilde{y})+\widetilde{H}_t(i+\tilde{y}))+\nu(\tilde{g}(\tilde{z})+\widetilde{H}_t(i+\tilde{z}))\}\\
&\geq \max_{0\leq\tilde{y}\leq\min\{l,N-i\},\ 0\leq\tilde{z}\leq\min\{l,N-j\}}\{\lambda(\tilde{g}(\tilde{y})+\widetilde{H}_t(i+\tilde{y}))+\nu(\tilde{g}(\tilde{z})+\widetilde{H}_t(i+\tilde{z}))\}\\
&= \lambda\max_{0\leq\tilde{y}\leq\min\{l,N-i\}}\{\tilde{g}(\tilde{y})+\widetilde{H}_t(i+\tilde{y})\}+\nu\min_{0\leq\tilde{z}\leq\min\{l,N-j\}}\{\tilde{g}(\tilde{z})+\widetilde{H}_t(i+\tilde{z})\}\\
&= \lambda\max_{0\leq y\leq\min\{l,N-i\}}\{g(y)+H_t(i+y)\}+\nu\min_{0\leq z\leq\min\{l,N-j\}}\{g(z)+H_t(i+z)\}\\
&= \lambda u_t(i,l,w_l)+\nu u_t(j,l,w_l),
\end{aligned}
$$

hence, it follows that $u_t(i,l,w_l)$ is concave in i.

Appendix C

Proof of (4.7) and (4.8) Since $I(w\in D(x))=I(w\in Z(x))-I(w\in Z(x+1))$, from (2.25) and (2.26) we have

$$
\begin{aligned}
R(i,\mathcal{D}(i)) &= \textstyle\sum_{x=1}^{k}\int r(w,x)(I(w\in Z(x))-I(w\in Z(x+1))dF(w)\\
&= \textstyle\sum_{x=1}^{k}\int r(w,x)I(w\in Z(x))dF(w)-\sum_{x=2}^{k+1}\int r(w,x-1)I(w\in Z(x))dF(w)\\
&= \textstyle\int r(w,1)I(w\in Z(1))dF(w)-\int r(w,k)I(w\in Z(k+1))dF(w)\\
&\qquad\qquad\textstyle+\sum_{x=2}^{k}\int \Delta r(w,x)I(w\in Z(x))dF(w)\\
&= \textstyle\int r(w,1)dF(w)+\sum_{x=2}^{k}\int \Delta r(w,x)I(w\in Z(x))dF(w)\\
&= \textstyle\int r(w,1)dF(w)+\sum_{x=2}^{k}\int_{w\in Z(x)} \Delta r(w,x)dF(w),
\end{aligned}
$$

$$
\begin{aligned}
P(j|i,\mathcal{D}(i)) &= \textstyle\sum_{x=1}^{k}p(j|i,x)\int(I(w\in Z(x))-I(w\in Z(x+1))dF(w)\\
&= \textstyle\sum_{x=1}^{k}p(j|i,x)\int I(w\in Z(x))dF(w)-\sum_{x=2}^{k+1}p(j|i,x-1)\int(w\in Z(x))dF(w)\\
&= \textstyle p(j|i,1)\int I(w\in Z(1))dF(w)-p(j|i,k)\int I(w\in Z(k+1))dF(w)\\
&\qquad\qquad\textstyle+\sum_{x=2}^{k}\Delta p(j|i,x)\int I(w\in Z(x))dF(w)\\
&= \textstyle p(j|i,1)+\sum_{x=2}^{k}\Delta p(j|i,x)\int_{w\in Z(x)}dF(w).
\end{aligned}
$$

Proof of (4.63) and (4.64) Since $I(w\in D(1))=1-I(w\in Z)$, $I(w\in D(x))=0$ for $1<x<k$, and $I(w\in D(k))=I(w\in Z)$, from (2.25) and (2.26) we have

$$
\begin{aligned}
R(i,\mathcal{D}(i)) &= \textstyle\int r(w,1)I(w\in D(1))dF(w)+\int r(w,k)I(w\in D(k))dF(w)\\
&= \textstyle\int r(w,1)(1-I(w\in Z))dF(w)+\int r(w,k)I(w\in Z)dF(w)\\
&= \textstyle\int r(w,1)dF(w)+\int\widetilde{\Delta}r(w)I(w\in Z)dF(w)\\
&= \textstyle\int r(w,1)dF(w)+\int_{w\in Z}\widetilde{\Delta}r(w)dF(w)
\end{aligned}
$$

$$
\begin{aligned}
P(j|i,\mathcal{D}(i)) &= \textstyle p(j|i,1)\int I(w\in D(1))dF(w)+p(j|i,k)\int I(w\in D(k))dF(w)\\
&= \textstyle p(j|i,1)\int(1-I(w\in Z))dF(w)+p(j|i,k)\int(w\in Z)dF(w)
\end{aligned}
$$

$$= p(j|i,1) + \tilde{\Delta}p(j|i) \int I(w \in Z)dF(w)$$
$$= p(j|i,1) + \tilde{\Delta}p(j|i) \int_{w \in Z} dF(w).$$

References

[1] Breiman, L.: Stopping-Rule Problems, *Applied Combinatorial Mathematics*, ed. E.F. Beckenbach, Chap. 10, 284-319 (1964)

[2] Burdett, K., and Malueg D.A.: The Theory of Search for Several Goods, *Journal of Economic Theory*, 24, 362-376 (1981)

[3] Chow, Y.S., Robbins, H., and Siegmund,D.: Great Expectations: The Theory of Optimal Stopping, *Houghton Mifflin Company*, Boston (1971)

[4] Carlson, J.A.: Joint Search for Several Goods, *Journal of Economic Theory*, 32, 337-345 (1984)

[5] Derman, C., Lieberman, G.J., and Ross, S.M.: A Sequential Stochastic Assignment Problem *Management Science*, 18, 7, March, 349-355 (1972)

[6] Fabian, T., Fisher, J.L., Sasieni, M.W., and Yardeni, A: Purchasing Raw Material on a Fluctuating Market, *Operations Research*, 7, 1, 107-122 (1959)

[7] Gilbert, J.P., and Mosteller, F: Recognizing the Maximum of a Sequence, *Journal of the American Statistic Association*, 61, 35-73 (1966)

[8] Hayes, R.H.: Optimal Strategies for Divestiture *Operations Research*, 17, 2, 292-310 (1967)

[9] Hochman, E.: An Optimal Stopping Problem of a Growing Inventory, *Management Science*, 19, 11, July, 1289-1291 (1973)

[10] Ikuta, S.: Basic Reserch on Optimal Customers Selection Problem, *Doctoral Dissertation*, Keio University, (1975)

[11] Ikuta, S.: Optimal Stopping Problem with Uncertain Recall: *Journal of the Operations Research Society of Japan*, 31, 2, June, 145-170 (1988)

[12] Ikuta, S.: The Optimal Stopping Problem in Which the Sum of the Accepted Offer's Value and the Remaining Search Budget is an Objective Function: *Journal of the Operations Research Society of Japan*, 35, 2, June, 172-193 (1992)

[13] Karlin, S.: Stochastic Models and Optimal Policy for Selling an Asset, *Studies in applied probability and management science*, Chap. 9, 148-158, Stanford University Press, 1962

[14] Kingsman, B.G.: Commodity Purchasing, *Operational Research Quarterly*, 20, 59-76 (1969)

[15] Kohn, M.G., and Shavell, S: The Theory of Search, *Journal of Economic Theory*, 9, 93-123 (1974)

[16] Karni, E., and Schwartz, A: Search Theory: The Case of Search with Uncertain Recall, *Journal of Economic Theory*, 16, 38-52 (1977)

[17] Lippman, S.A., and McCall, J.J.: The Economics of Job Search: A Survey, *Economic Inquiry*, 14, June, 155-189 (1976)

[18] Lippman, S.A., and McCall, J.J.: Job Search in a Dynamic Economy, *Journal of Economic Theory*, 12, 365-390 (1976)

[19] Landsberger, M., and Peled, D.: Duration of Offers, Price Structure, and the Gain from Search, *Journal of Economic Theory*, 16, 17- 37 (1977)

[20] Morris, W.T.: Some Analysis of Purchasing Policy, *Management Science*, 5, 4, July, 443-452 (1959)

[21] MacQueen, J., and Miller Jr, R.G.: Optimal Persistence Policies, *Operations Research*, 8, 3, 362-380 (1960)

[22] McCall, J.J.: The Economics of Information and Optimal Stopping Rules, *Journal of Business*, July, 300-317 (1965)

[23] Miller, B.L.: A Queing Reward System with Several Customer Class, *Management Science*, 16, 3,(1969)

[24] Mastran, D.V., and Thomas, C.J.: Decision Rules for Attacking Targets of Opportunity, *Naval Research Logistics Quarterly*, 20, 661-672 (1973)

[25] Magirou, V.F.: Stockpiling under Price Uncertainty and Storage Capacity Constraints, *European Journal of Operational Research*, 11, 233-246 (1982)

[26] Morgan, P., and Manning, R.: Optimal Search, *Econometrica*, 53, 4, July, 923-944 (1985)

[27] Rosenfield, D.B.: Optimal Adaptive Price Search, *Journal of Economic Theory*, 25, 1-20 (1981)

[28] Rosenfield, D.B., Shapiro, R.D., and Butler, D.A.: Optimal Strategies for Selling an Asset, *Management Science*, 29, 9, September, 1051-1061 (1983)
[29] Rothschild, M.: Searching for the Lowest Price When the Distribution of Prices Is Unknown, *Journal of Political Economy*, 82, 4, 689-711
[30] Sakaguchi, M.: Dynamic Programming of Some Sequential Sampling Design, *Journal of Mathematical Analysis and Applications*, 2, 446-466 (1961)
[31] Sakaguchi, M. An Investment Problem: An Optimal Stopping Problem in Which Two Stops Are Required, *Journal of the Operations Research Society of Japan*, 15, 1, March, 45-52 (1972)
[32] Sargent, T.J.: Search, *Dynamic Macroeconomic Theory*, 57-91, Harvard University Press, (1987)
[33] Taylor, H.M.: Evaluating a Call Option and Optimal Timing Strategy in the Stock Market, *Management Science*, 14, 1, September, 111-120 (1967)
[34] Weitzman, M.L.: Optimal Search for the Best Alternative, *Econometrica*, 47, 3, May, 641-654 (1979)
[35] Yanai, H.: On a Class of Optimal Stopping Rule Problems, *Journal of the Operations Research of Japan*, 8, 2, February, 66-90 (1966)
[36] Zuckerman, D.: On Preserving the Reservation Wage Property in a Continuous Job Search Model, *Journal of Economic Theory*, 34, 175-179 (1984)

Seizo IKUTA: Institute of Socio–Economic Planning,
University of Tsukuba, Tsukuba, baraki, 305, Japan