Journal of Informatics and Mathematical Sciences
Volume 2 (2010), Numbers 2 & 3, pp. 121–139

http://www.rgnpublications.com

A Parametric Statistics as A Supporting Tool to the Social Sciences

John E. Goulionis and Dimitrios I. Stengos

Abstract. *Partially Observable Markov Decision Processes* (POMDPs) have recently been suggested as a suitable model to formalizing the planning of educational management. In this paper, we discuss a specialization of POMDPs that is tailored to a frequently re-occurring type of educational management problem. Furthermore under some reasonable conditions it is shown that there exists an optimal policy in the class of control limit policies and we develop a solution procedure utilizing these properties of the specialized form.

1. Introduction

A *Partially Observable Markov Decision Process* (POMDP) is a general sequential decision-making model, where the effects of actions are nondeterministic and only partial information about world states is available. The average cost criterion is a popular criterion for optimization of stochastic dynamical systems over an infinite time horizon. On the theoretical side Blackwell [3] considered the *discounted cost* (DC) criterion in great details. The relation between the discounted and average case becomes apparent via Tauberian theorems, see Arapostathis [1].

Cavazos-Cadena [5], showed that the existence of bounded solutions to the *average cost optimality equation* (ACOE) necessarily impose a very strong recurrence structure on the model. Recently POMDPs have been suggested as providing a suitable, integrated approach to this type of educational management problems see Sondik [17] and Goulionis [6]. Unfortunately, the computational burden associated with solving POMDPs is overwhelming, recluding their application to problems of practical size Papadimitriou [14].

However,for many specialized problems, the full-blown generality of the POMDP approach and its associated solution methods is superfluous. As our educational systems increases in size and complexity, and as they become

2000 *Mathematics Subject Classification.* 90C40, 90B50, 90C39, 90C15, 93E20, 60J70.
Key words and phrases. Statistics in education; Average cost Markov processes; Dynamic programming.

increasingly dependent upon the devices and techniques of the new educational technology, a systematic quantitative approach to the design and operation of these educational systems is a vital necessity. For this reason we begin the process for the simple yet very important system composed of a class of students.

In this paper, we apply partially observable Markov decision processes to an educational system with two available actions (educational-methods). The first method is cheap, and the second method is luxurious supported with computers and presentations and we have partially observed case. Our results were presented by means of a simple problem, but the main ideas can be readily put to use in many applications.

The paper is organized as follows. In section 2, the model is described in detail.

In section 3 an optimal replacement problem is formulated as POMDP and we determine the structural properties of optimal policies for the average cost criterion. In section 4 we illustrate these results in the context of machine replacement problem with two states.

2. **Model Description and Assumptions**

In the modeling of physical systems the concept of the state of the physical system has proved to be a very valuable tool for the characterization of system performance Sondik [17], Goulionis [5], [7]. This idea may also be an important aid to the description of the learning process.

Thus in the class of the students the internal state of a class is measured of the internal state of the students that constitute the class. We shall use the internal state of a student as a representation of his learning characteristics. The internal state of a student depends on different factors, as hereditary roots, familial and social environments, personal model of thinking, preexisted knowledge, sentimental reasons and generally psychological factors. Therefore, the internal state of a class depends on many factors and for this reason is unknown. However we can take a sense of this internal state by some observations, for example (score in a test, participation in the learning process, the language of a body etc) see Goulionis [8].

In this section we briefly describe the *partially observable Markov decision processes* (POMDPs) model decision theoretic planning problems in which an agent must take a sequence of decisions to minimize its utility given uncertainty in the effects of its actions and its current state.

A POMDP model is a tuple $(X, D, P, R, \Theta, \beta)$. The partially observable Markov decision process consists of a core process, an observation process and a decision process. X is a finite set of stochastic variables. In this paper we have two states $X = \{1, 2\}$. Let $\{x_t, t = 0, 1, \ldots\}$ denotes the state of the class at time t. x_t takes values in $X \equiv \{1, 2\} \equiv \{\text{good}, \text{bad}\}$. The state of the class is only partially observed and we have two actions (educational-methods) available in order to control

the situation of the class. The first method is cheap, and the second method is luxurious supported with computers and presentations. These methods are coded in $0,1$ respectively; the set of actions at each time are $D \equiv \{0,1\}$. We denote by $\{u_t,\ t = 0,1,2,\ldots\}$ the control process. The value of u_t denotes the decisions taken at time t. The teacher is unable to observe the state of the class directly and must make his/her decisions sequentially based upon partial information time. State transitions occur according to a Markov Chain whose transition probabilities are determined by the choice of the material to be presented to the class. To accomplish the effect of the teaching method upon the internal state knowledge of a class by transitions from one state to another state, we have a transition probability matrix. The state process evolves according to the transition probabilities $P_{x_t x_{t+1}}(u_t)$ define by

$$p_{ij}(a) = P\{x_{t+1} = j | x_t = i, u_t = a\}, \text{ where } i,j = 1,2, a \in D,\ t = 0,1,\ldots .$$

The transition probability matrices $P(u_t)$, $u_t \in D$, with entries $P_{x_t x_{t+1}}(u_t)$, are given by:

$$P(0) = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}, \qquad P(1) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}. \tag{2.1}$$

The observation process is related to the state and the control processes by means of the conditional probabilities $r_{x_t, y_{t+1}}(u_t)$ defined by

$$r_{i\theta}(a) = P\{y_{t+1} = \theta | x_t = i, u_t = a\}, \quad i = 1,2$$

with $r_{x_t, y_{t+1}}(u_t)$ the entries of the observation matrices $R(u_t)$, $u_t \in D$, given by:

$$R(0) = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}, \qquad R(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.2}$$

where $r_{11}, r_{12} \in [0.5, 1.0]$ is the probability of making a correct observation.

The signals for our model represent outcomes of the tests. For simplicity we consider two types of observations are coded in $1,2$. We consider that the observation of type 1 ($\theta = 1$), is favorable for the state 1 (good state of a class), while the observation of type 2 ($\theta = 2$),is favorable for the state 2, (bad state of a class). For example we consider that we have an observation of type 1 ($\theta = 1$), if we have success in the test over 60%. Thus the sets of signals is $\Theta = \{1,2\}$. Since the true state of the class is not know, all the relevant information for selecting the control action at time t is summarized by Bertsekas [2], [3] the conditional probability distribution (also referred to as the information vector or the sufficient statistic)

$$\pi(t) = (\pi_1(t), \pi_2(t)) = (1 - p(t), p(t)), \tag{2.3}$$

where $p(t)$ is the probability that the class is in the bad state at time t given past observations and actions, $0 \leq p(t) \leq 1$. However, to do this we first need to supply priors on all hidden variables at time $t = 0$. We note that the current prior can be replaced by a more flexible model that targets different classes of students and exploits other context information. There are logistic regression models developed for this purpose [1]. In this paper the initial probabilities $\pi(0) = (P\{x_0 = 1\}, P\{x_0 = 2\})$ are assumed given. That is, $p(t)$ is defined as:

$$p(t) = P\{x_t = 2 | y_t, \ldots, y_1, u_{t-1}, \ldots, u_0\}, \quad t = 1, 2, \ldots$$

$$p(0) \equiv \pi_2(0).$$

Initial belief state. When the model of the dynamics has been defined, we can expand it as many time steps as needed and use it to compute belief updates. However, to do this we first need to supply priors on all hidden variables at time $t = 0$. There are logistic regression models developed for this purpose Goulionis [7]. We will often denote $\pi(t)$ and $p(t)$ by π and p respectively, omitting explicit dependence on t. We set $C^a \equiv (c(1,a), c(2,a))$, $a \in D$, where $c(i,\alpha)$ is the scalar valued cost accrued, when the current state is $i \in S$ and action is $u_t = \alpha \in D$, $t = 0, 1, 2, \ldots$.

In this model we assume that $c(1,0) = c_1$, $c(2,0) = c_2$, $c(1,1) = c(2,1) = R$, where $c_1 < c_2 < R$.

The objective is to find, an optimal policy among the admissible policies, such that it minimizes a given performance index, typically the expected discounted cost or the expected long-run average cost. These costs are defined in terms of the state x_t by:

Discounted-cost (DC).

$$J_\beta(\delta, \pi(0)) := \lim_{n\to\infty} E^{\delta}_{\pi(0)}\left[\sum_{t=0}^{n} C(x_t, u_t)\right], \quad 0 < \beta < 1\,. \tag{2.4}$$

Average-cost (AC).

$$J(\delta, \pi(0)) := \lim_{n\to\infty} \sup \frac{1}{n} E^{\delta}_{\pi(0)}\left[\sum_{t=0}^{n-1} C(x_t, u_t)\right] \tag{2.5}$$

respectively, and in terms of the information vector $\pi(t)$ by:

$$J_\beta(\delta, \pi(0)) := \lim_{n\to\infty} E^{\delta}_{\pi(0)}\left[\sum_{t=0}^{n} \beta^t \pi(t) C^{u_t}\right], \tag{2.6}$$

and

$$J(\delta, \pi(0)) := \lim_{n\to\infty} \sup \frac{1}{n} E^{\delta}_{\pi(0)}\left[\sum_{t=0}^{n-1} \pi(t) C^{u_t}\right]. \tag{2.7}$$

The equivalence, in the sense of equal optimal costs for each $\pi(0) \in \Pi$, of the optimization problems defined using criteria (2.4) and (2.5), and similarly for problems specified using (2.6) and (2.7), is shown in Sawaragi [14].

Since the state of the machine is not known at time t, we will work with $J_\beta(\delta,\cdot)$ and $J(\delta,\cdot)$. We define

$$J(\pi) \equiv \inf_{\delta} J_\beta(\delta,\pi).$$

Then, $V_\beta(\pi)$is the total expected discounted cost accrued when an optimal policy is selected, given that the initial information vector is π, and future costs are discounted at rate β. It is well known Bertsekas [4] and Ross [13] that $V_\beta(\pi)$ is the unique solution of:

$$V_\beta(\pi) = \min_a \left\{ \pi\, C^a + \beta \sum_\theta \{\theta/\pi,\alpha\} V_\beta(\pi/\theta,\alpha) \right\}, \tag{2.8}$$

$$\{\theta/\pi,a\} = \pi\, P^a R^a_\theta\, e, \tag{2.9}$$

where $\{\theta/\pi,a\} = \pi\, P^a R^a_\theta\, e$ is the probability that the next observation will be θ, given the probability distribution, π and action a, with $e \equiv (1,1)'$ and R^a_θ the 2×2 diagonal matrix with entries $r_{i\theta}(a)$, $i = 1,2$. $T(\pi,\theta,a)$ is the updated conditional probability given observation θ, action a and prior distribution π, and using Bayes' rule it is given by

$$T(\pi,\theta,a) = \frac{\pi\, P^a R^a_\theta}{\{\theta,\pi,a\}}\,. \tag{2.10}$$

When computing optimal policies in the infinite horizon case, we need only consider stationary policies see Bertsekas [4]. A stationary policy is denoted by $(\delta)^\infty = (\delta,\delta,\cdots)$. Similarly, define:

Optimal-average cost.

$$g \equiv J(\pi) \equiv \inf_{\delta} J(\delta,\pi), \quad \pi \in \Pi.$$

Then, g is the expected optimal average cost, and it satisfies the functional equation:

$$g + h(\pi) = \min_a \left\{ \pi C^\alpha + \sum_\theta \{\theta/\pi,\alpha\} h(T(\pi,\theta,\alpha)) \right\} \ \forall\ \pi \in \Pi. \ \text{(ACOE)} \tag{2.11}$$

(Conditions for the existence of a constant g and measurable map h, satisfying the average cost optimality equation (2.10) for the problem considered here, are given in Fernandez [1]). The key points are:

(i) the observation and control spaces, Θ and A respectively, are finite; and
(ii) the costs $C(x(t),a(t))$, $x(t) \in X$, $a(t) \in A$, $t = 0,1,2,\ldots$, are uniformly bounded.

3. **Optimal-replacement Problem (Average-cost)**

The average cost criterion (equivalently, the long-run average cost) is a popular criterion for optimization of stochastic dynamical systems over an infinitive time horizon. We wish to determine the structural properties of optimal policies for the average cost criterion.

Theorem 3.1. *If there exists a real bounded function h, $h:\Pi\to R$ and a constant g, such that:*

$$g+h(\pi)=\min_{a}\left\{\pi C^{\alpha}+\sum_{\theta}\{\theta/\pi,\alpha\}h(T(\pi,\theta,\alpha))\right\},\ \forall\ \pi\in\Pi.\quad\text{(ACOE)}$$

Then, it can be shown that g is the optimal average cost, and that any stationary policy δ^* attaining the minimum above is average cost optimal.

$$g=J(\delta^*,\pi)=J(\pi),\quad \forall\ \pi\in\Pi.\qquad\text{(by Ross [13])}$$

Let now, $\pi,\pi^*\in\Pi$ and $h_\beta(\pi):=V_\beta(\pi)-V_\beta(\pi^*)$, $g_\beta=(1-\beta)$. $V_\beta(\pi^*)\ \forall\ \pi\in\Pi$, $(0\le\beta<1)$, then the average cost optimality equation (ACOE) takes the following form:

$$g_\beta+h_\beta(\pi)=\min_{a}\left\{\pi q^{\alpha}+\sum_{\theta}\{\theta/\pi,\alpha\}h_\beta(T(\pi,\theta,\alpha))\right\}\ \forall\ \pi\in\Pi.$$

It can be shown, Fernandez [1] that a necessary condition for the existence of a bounded solution to the (ACOE) is that the following boundedness condition holds.

Uniform-boundedness. There is a constant $k\succ 0$ such that:

$$|V_\beta(\pi)-V_\beta(\pi^*)|\le k\quad\forall\ \beta,(0\le\beta<1)\ \text{ and }\ \forall\ \pi,\pi^*\in\Pi$$

Let $\pi_0\in\Pi$. We consider the set:

$$S(\pi_0)=\bigcup_{t=0}^{\infty}S_t(\pi_0),$$

$$S_0(\pi_0)=(\pi_0),$$

$$S_t(\pi_0)=\{T(\pi,\theta,\alpha):\pi\in S_{t-1}(\pi_0),\theta\in M,\ a\in A\},\quad t\ge 1.$$

$S(\pi_0)$ is countable set, since the countable union of countable sets is itself countable.

Theorem 3.2. *If there exists a constant $0<k<\infty$ such that:*

$$h_\beta(\pi)\equiv|V_\beta(\pi)-V_\beta(\pi^*)|\le k\ \ \forall\ \ \pi\in S(\pi_0)\ \ (0\prec\beta<1)\ \ (\textit{Uniform-boundeness})$$

then:

(i) *there exist, a bounded function* $h(\pi)$ *with* $\pi \in S(\pi_0)$*, a constant* g_{π_0}*, and some sequence* $\{\beta_n\}$*,* $\beta_n \in (0,1)$*,* $\beta_n \to 1$ $(n \to \infty)$ *such that:*

$$h_{\beta n}(\pi) \to h(\pi) \quad \forall\ \pi \in S(\pi_0) \ \text{ and } (n \to \infty)$$

$$g_{\beta n} \to g_{\pi_0}\ (n \to \infty)$$

(ii) *The constant* g_{π_0}*, and function* $h(\pi)$ *satisfy the average cost optimality equation.*

$$g_{\pi_0} + h(\pi) = \min_a \left\{ \pi\, q^\alpha + \sum_\theta \{\theta/\pi, \alpha\} h(T(\pi, \theta, \alpha)) \right\} \quad \forall\ \pi \in S(\pi_0).$$

(iii) $g_{\pi_0} = J(\pi)\ \forall\ \pi \in S(\pi_0)$.

(iv) $g_{\pi_0} = J(\pi)$. *From Theorem* 3.1.

Proof. Fernandez [1]. □

Theorem 3.3. *Under the assumption that* $h_\beta(\pi) \equiv |V_\beta(\pi) - V_\beta(\pi^*)| \leq k$*,* $\forall$ $\pi \in S(\pi_0)$ $(0 \prec \beta < 1)$ (*Uniform-boundedness*)*, then there exist a bounded function* $h(\pi)$ *with* $\pi \in \Pi$ *and a constant* g*, such that:*

$$g + h(\pi) = \min_a \left\{ \pi\, C^0 + \sum_\theta \{\theta/\pi\} h(T(\pi, \theta)), \pi\, C^1 \right\}, \tag{3.1}$$

because $h(e_1) = 0$ *and* $g = J(\pi)$*,* $\forall\ \pi \in \Pi$.

Proof. We consider the set

$$S_0(\pi) = \{\pi\},$$

$$S_t(\pi) = \{T(\pi, \theta) : \pi \in S_{t-1}(\pi), \theta \in M\} \cup \{e_1\}, \quad t \geq 1,$$

$$S(\pi) = \bigcup_{t=0}^{\infty} S_t(\pi),$$

is countable set, since the countable union of countable sets is itself countable. From theorems (3.2), (3.3) there exists, a bounded function $h(\pi)$ with $\pi \in \Pi$, and a constant g, such that:

$$g + h(\pi) = \min_a \left\{ \pi\, C^0 + \sum_\theta \{\theta/\pi\} h(T(\pi, \theta)), \pi\, C^1 \right\}.$$

It is valid that: $S(e_1) \subseteq S(\pi)\ \forall\ \pi \in \Pi$, and therefore $g_\pi = J(e_1)$, $\forall\ \pi \in \Pi$, hence g_π independent of π. □

Proposition 3.4. *If a function* $h_\beta(p)$ *is uniformly bounded. Then for a fixed* $p \in [0,1]$*, it is average-cost-optimal to take action "*α*" at* p*, denoted as* $\delta^*(p) = \alpha$*, if there is a sequence* $\{\beta_n\} \subseteq (0,1)$*, with* $\beta_n \uparrow 1$*, such that it is* β_n*-discount optimal to take action "*α*" at* p*, denoted as* $\delta^*_{\beta_n}(p) = \alpha$.

Proof. Fernandez [1]. □

4. The Two-state Replacement Problem

We wish to determine the structural properties of optimal policies for the average cost criterion by examining (3.1). We restrict the state space of the core process and the observation process to a two state case (1-good state and 2-bad state). Hence we can write $\pi = (1-p, p)$. Here, p, is interpreted as a priori probability of the system being failed. At each time period, the state of the system is monitored incompletely by some monitoring mechanism. The outcome of the monitoring is classified into finite levels $\Theta = \{1, 2\}$.

$T(p,\theta) \equiv T_2(\pi,\theta)$ is a posteriori conditional probability of the core state being in the bad state, given decision $\theta = 0$ was made, observation θ obtained, and an a priori probability, p, of the system being failed. It follows from (2.9) and (2.10) that:

$$T(p,\theta) = \frac{a_\theta + \beta_\theta p}{\gamma_\theta + \delta_\theta p}, \quad 0 \le p \le 1, \ \theta = 1,2, \tag{4.1}$$

$$\{\theta/p\} = \gamma_\theta + \delta_\theta p, \qquad 0 \le p \le 1, \ \theta = 1,2, \tag{4.2}$$

$$\alpha_\theta = p_{12} r_{2\theta} > 0, \quad \beta_\beta = (p_{22} - p_{12}) r_{2\theta}, \ \gamma_\theta = p_{11} r_{1\theta} + p_{12} r_{2\theta} > 0 \text{ and}$$

$$\alpha_\theta = p_{21} r_{1\theta} + p_{22} r_{2\theta} - p_{11} r_{1\theta} - p_{12} r_{2\theta} = -|P| r_{1\theta} + |P| r_{2\theta}, \quad \theta = 1,2\,. \tag{4.3}$$

The next lemma collects some of the properties of the maps $T(p,1)$, $T(p,2)$, $0 \le p \le 1$.

Lemma 4.1. *Let $0 \le p \le 1$, $r_{11}, r_{22} \in (0.5, 1)$, the following holds:*

(i) *The functions $T(p,1)$, $T(p,2)$, are monotone nondecreasing for each $p \in (0,1)$.*

(ii) *The function $T(p,1)$, $0 \le p \le 1$ is convex and the function $T(p,2)$, $0 \le p \le 1$ is concave.*

(iii) $T(p,1) \prec T(p,2)$, $0 \le p \le 1$

(iv) $T(1,\theta) \prec 1 \ \ (\theta = 1,2)$

(v) *The function $T(p,1)$ has single fixed point $\xi_1 \in (0,1)$, $T(\xi_1,1) = \xi_1$ and $p \prec T(p,1) \prec \xi_1 \ \forall \ p \in [0,\xi_1)$ and $\xi_1 \prec T(p,1) \prec p \ \forall \ p \in (\xi_1, 1]$.*

(vi) *The function $T(p,2)$ has single fixed point $\xi_2 \in (0,1)$, $T(\xi_2,1) = \xi_2$ and $p \prec T(p,2) \prec \xi_2 \ \forall \ p \in [0,\xi_2)$ and $\xi_2 \prec T(p,2) \prec p \ \forall \ p \in (\xi_2, 1]$.*

(vii) $\xi_1 \prec \xi_2$.

Proof. The properties of the maps $T(p,1)$, $T(p,2)$ follow by simple algebraic operations on the expressions for the given quantities. □

Now, concerning the quantities defined above, the following assumptions are made.

(A-1) The 2×2 transition probability matrix P is totally positive of order $2\,(TP_2)$,that is,

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \geq 0\,.$$

(A-2) The 2×2 probability matrix R is totally positive of order $2\,(TP_2)$, that is,

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \geq 0.$$

(A-3) $p_{12} \succ 0$.

(A-4) $c_1 \prec c_2 \prec R$.

Now, let $V_\beta(\pi)$ denote the optimal expected total discounted cost over an infinite horizon with an initial state π. Then $V_\beta(\pi)$ satisfies the following recursion:

$$V_\beta(\pi) = \min\left\{\pi\, C^0 + \beta \sum_\theta \{\theta/\pi\} V_\beta(T(\pi,\theta)), \pi\, C^1 + \beta\, V_\beta(e_1)\right\}$$

$$\text{for } \pi \in R^2,\ \pi = (1-p, p)$$

or

$$V_\beta(p) = \min\left\{c_1 + c\,p + \beta \sum_{\theta=1}^{2} \{\theta/p\}\, V_\beta(T(p,\theta)), R + V_\beta(0)\right\},$$

where $c = c_2 - c_1$ ($\succ 0$).

Theorem 4.2. *The function $h_\beta(\pi) \equiv |V_\beta(\pi) - V_\beta(e_1)|$ is uniformly bounded.*

Proof.

$$V_\beta(\pi) = \min\begin{cases} \pi C^0 + \beta \sum\limits_{\theta\in\Theta} \{\theta|\pi\} V_\beta(T(\pi,\theta)) \\ \pi C^1 + \beta V_\beta(e_1) \end{cases} \qquad \text{for } \pi \in R^2,\ e_1 = (1,0)\,.$$

For $\pi \in R^2$, $\pi = (1-\pi_2, \pi_2)$, $e_1 = (1,0)$, where $(0 \leq \beta < 1)$, we have:

$$V_\beta(\pi) \leq \pi\, C^1 + \beta.\ \ V_\beta(e_1) \leq R + V_\beta(e_1).$$

Therefore:

$$0 \prec V_\beta(\pi) - V_\beta(e_1) \prec R,$$

and the function $h_\beta(\pi) \equiv |V_\beta(\pi) - V_\beta(e_1)|$ is uniformly bounded. □

Proposition 4.3. (i) *The optimal β-discounted cost function $V_\beta(p)$, $0 \leq p \leq 1$ is concave and nondecreasing.*

(ii) $0 \prec V_\beta(p) \leq \dfrac{R}{1-\beta}$, $0 \leq p \leq 1$.

Proof. (i) The function $V_\beta(\pi)$, $\pi \in \Pi$ is increasing and concave; see Astrom [2].

$$\text{(ii)} \quad V(p) = \min\left\{c_1 + c\,p + \beta \sum_{\theta=1}^{2} \{\theta/p\}\, V_\beta(T(p,\theta)), R + \beta\, V_\beta(0)\right\}, \quad 0 \le p \le 1 \tag{4.4}$$

From above optimality equation

$$V_\beta(p) \le R + \beta V_\beta(0), \quad 0 \le p \le 1\,.$$

For $p = 0$ we take $V_\beta(0) \le R + \beta V_\beta(0)$.

Hence, $V_\beta(0) \le \dfrac{R}{1-\beta}$.

Therefore,

$$V_\beta(p) \le R + \beta\,\frac{R}{1-\beta} = \frac{R}{1-\beta}, \quad 0 \le p \le 1\,. \tag{4.5}$$

□

Define $W_\beta(p)$, $0 \le p \le 1$, the expected discounted cost to be accrued using the policy $\delta(p) = 0$, $0 \le p \le 1$.

$$W_\beta(p) = c_1 + p\,c + \beta \sum_{\theta} \{\theta/p\} W_\beta(T(p,\theta)), \quad 0 \le p \le 1, \tag{4.6}$$

where $c = c_2 - c_1$ ($\succ 0$).

For finding W_β we consider the functions $W_n(p)$, $n = 0, 1, 2, 3\ldots$, $0 \le p \le 1$ defined inductively as:

$$W_0(p) = 0, \quad 0 \le p \le 1$$

$$W_n(p) = c_1 + p\,c + \beta \sum_{\theta=1}^{2} \{\theta/p\} W_{n-1}(T(p,\theta)), \quad 0 \le p \le 1. \tag{4.7}$$

Lemma 4.4.

$$W_n(p) = (A_n\,p + B_n)\,c + \frac{1-\beta^n}{1-\beta}\,c_1\,, \quad n = 0, 1, 2, \ldots\,, \tag{4.8}$$

where

$$A_n = 1 + \beta |P| A_{n-1}, \quad B_n = B(p_{12} A_{n-1} + B_{n-1}), \quad n = 1, 2, \ldots$$

$$A_0 = B_0 = 0.$$

Proof. An induction argument shows that, $W_1(p) = c_1 + p\,c = c_1 + (A_1\,p + B_1)\,c$, $A_1 = 1$, $B_1 = 0$. We suppose that lemma is valid for some $n \ge 1$. For $0 \le p \le 1$ we have:

$$W_{n+1}(p) = \left[c_1 + p\,c + \beta \sum_{\theta} \{\theta/p\} W_n(T(p,\theta)\right]$$

$$= c_1 + p\,c + \beta \sum_{\theta} \{\theta/p\} \left[(A_n\, T(p,\theta) + B_n)\, c + \frac{1-\beta^n}{1-\beta}\, c_1 \right]$$

$$= c_1 + p\,c + \beta \left(A_n \sum_{\theta} \{\theta/p\} [T(p,\theta) + B_n)\, c + \beta\, \frac{1-\beta^n}{1-\beta} c_1 \right)$$

$$= \frac{1-\beta^{n+1}}{1-\beta}\, c_1 + c\,p + \beta \left(A_n \sum_{\theta} (\alpha_\theta + \beta_\theta\, p) + B_n \right) c$$

$$= \frac{1-\beta^{n+1}}{1-\beta}\, c_1 + c\,p + \beta\, \left[(A_n((a_1 + a_2) + (\beta_1 + \beta_2)p) + B_n) \right] c,$$

where

$$\alpha_1 + \alpha_2 = p_{12}\,(r_{21} + r_{22}) = p_{12}\,,$$

$$\beta_1 + \beta_2 = |P|\,(r_{21} + r_{22}) = |P|\,.$$

Hence,

$$W_{n+1}(p) = \frac{1-\beta^{n+1}}{1-\beta}\, c_1 + c\,p + \beta\, [(A_n(p_{12} + |P|\,p) + B_n)]\, c$$

$$= \frac{1-\beta^{n+1}}{1-\beta}\, c_1 + (1 + \beta\,|P| A_n)\, c\,p + \beta\,(p_{12} A_n + B_n)\, c$$

$$= \frac{1-\beta^{n+1}}{1-\beta}\, c_1 + (A_{n+1}\,p + B_{n+1})\, c\,.$$

Therefore the lemma is valid for $n+1$. □

The following result gives an explicit solution for W_β.

Proposition 4.5. *The function* $W_\beta(p)$, $0 \le p \le 1$

$$W_\beta(p) = \frac{1}{1-\beta}\, c_1 + \frac{(1-\beta)\,p + \beta\, p_{12}}{(1-\beta)(1-\beta|P|)}\, c\,, \quad 0 \le p \le 1.$$

Proof. Note $W_n \to W_\beta$ when $n \to \infty$.
Where,

$$W_0(p) = 0, \quad 0 \le p \le 1\,,$$

$$W_n(p) = c_1 + p\,c + \beta \sum_{\theta=1}^{2} \{\theta/p\} W_{n-1}(T(p,\theta)), \quad 0 \le p \le 1\,,$$

$$W_\beta(p) = (Ap + B)\, c + \frac{1}{1-\beta}\, c_1, \quad 0 \le p \le 1$$

follows from Lemma 4.4.

Where,

$$A = \lim_{n\to\infty} A_n\,, \quad B = \lim_{n\to\infty} B_n.$$

Limiting arguments produce the result,

$$A = 1 + \beta|P|A,$$

$$B = \beta(p_{12}A + B).$$

Hence,

$$A = \frac{1}{1-\beta|P|}, \quad B = \frac{\beta\, p_{12}}{(1-\beta)(1-\beta|P|)}$$

and finally

$$W_\beta(p) = \frac{1}{1-\beta}\, c_1 + \frac{(1-\beta)\, p + \beta\, p_{12}}{(1-\beta)(1-\beta|P|)}\, c, \quad 0 \le p \le 1\,. \tag{4.9}$$

□

Proposition 4.6. *The strategy $\delta^\infty$with control function $\delta(p) = 0,\ 0 \le p \le 1$ is β-optimal if and only if:*

$$c_1 + \frac{1+\beta\, p_{12}}{1-\beta\,|P|}\, c \le R. \tag{4.10}$$

Proof. The function W_β is β-optimal if and only if:

$$W_\beta(p) \le R + \beta\, W_\beta(0), \quad 0 \le p \le 1\,. \tag{4.11}$$

Then by proposition (4.5) it follows that:

$$W_\beta(p) = \frac{1}{1-\beta}\, c_1 + \frac{(1-\beta)\, p + \beta\, p_{12}}{(1-\beta)(1-\beta|P|)}\, c\,.$$

Thus we obtain that (4.11) it is valid if and only if:

$$c_1 + \frac{p+\beta\, p_{12}}{1-\beta\,|P|}\, c \le R, \quad 0 \le p \le 1\,. \tag{4.12}$$

But the above (4.12) it is valid if and only if (4.10) is valid. Therefore we conclude that the function W_β is β-optimal if and only if it is valid (4.10). Then the strategy $\delta(p) = 0,\ 0 \le p \le 1$ is β-optimal, if and only if it is valid (4.10). □

Proposition 4.7. *The strategy $\delta^\infty$with control function $\delta(p) = 0,\ 0 \le p \le 1$ is β-optimal $\forall\ \beta \in (0,1)$ if and only if:*

$$c_1 + \frac{1+p_{12}}{1-|P|}\, c \le R\,. \tag{4.13}$$

Proof. The function $f(\beta) := \dfrac{1+\beta\, p_{12}}{1-\beta|P|}\, c$, $0 \prec \beta \prec 1$ is monotone increasing and

$$\lim_{\beta \to 1^-} f(\beta) = \frac{1+p_{12}}{1-|P|}\, c\,.$$

Note that $|P| = p_{22} - p_{12} \prec 1$, because $p_{12} \succ 0$ (Assumption A-3)).

Therefore the condition (4.13) follows from Proposition 4.6 with ($\beta = 1$). □

Then by Theorems 3.4 and 4.2 it follows, that there exist, a bounded function $h(p)$, $0 \le p \le 1$ and a constant g such that,

$$g + h(p) = \min_{a}\left\{c_1 + c\,p + \sum_{\theta}\{\theta/p\}\,h(T(p,\theta)), R\right\}, \quad 0 \le p \le 1\,, \quad (4.14)$$

$$h(0) = 0\,,$$

$$g = \lim_{\beta\to 1^-}(1-\beta)\,V_\beta(0)\,. \quad (4.15)$$

The following proposition provides necessary and sufficient condition, such that the strategy "not replace the system" to be average cost AC-optimal.

Proposition 4.8. *The strategy δ^∞ with control function $\delta(p) = 0$, $0 \le p \le 1$ is average cost (AC) optimal if and only if:*

$$c_1 + \frac{1+p_{12}}{1-|P|}\,c \le R \quad (4.16)$$

and then the average cost accrued by this policy is

$$g = c_1 + \frac{p_{12}}{1-|P|}\,c\,.$$

Proof. Suppose that (4.12) holds. Then by proposition (4.7) the strategy δ^∞ with $\beta(p) = 0$, $0 \le p \le 1$ is β-optimal for each $\beta \in (0,1)$ and

$$V_\beta(p) = W_\beta(p), \quad 0 \le p \le 1\,. \quad (4.17)$$

We have from (4.9), (4.15) and (4.17) by simple algebraic operations,

$$\begin{aligned} g &= \lim_{\beta\to 1^-}(1-\beta)\,V_\beta(0) \\ &= \lim_{\beta\to 1^-}(1-\beta)\,W_\beta(0) \\ &= \lim_{\beta\to 1^-}\left(c_1 + \frac{\beta\,p_{12}}{1-\beta|P|}\,c\right) \\ &= c_1 + \frac{p_{12}}{1-|P|}\,c\,. \end{aligned} \quad (4.18)$$

Also for each $\beta \in (0,1)$ we have:

$$\begin{aligned} h_\beta(p) &= V_\beta(p) - V_\beta(0) \\ &= W_\beta(p) - W_\beta(0) \\ &= \frac{c}{1-\beta|P|}\,p, \quad 0 \le p \le 1\,. \end{aligned}$$

For $\beta \to 1^-$ we take:

$$h(p) := \lim_{\beta\to 1^-} h_\beta(p) = \frac{c}{1-|P|}\,p\,, \quad 0 \le p \le 1\,.$$

Now we prove that g and $h(p)$, $0 \le p \le 1$ satisfy the optimization equation (4.14). We have

$$\begin{aligned} & c_1 + c\,p + \sum_{\theta} \{\theta/p\}\, h(T(p,\theta)) \\ & \quad = c_1 + c\,p + \frac{c}{1-|P|} \sum_{\theta} \{\theta/p\}\, T(p,\theta)) \\ & \quad = c_1 + c\,p + \frac{c}{1-|P|} \sum_{\theta} (\alpha_\theta + \beta_\theta\, p) \\ & \quad = c_1 + c\,p + \frac{c}{1-|P|} (a_1 + a_2 + (\beta_1 + \beta_2)\, p) \\ & \quad = c_1 + c\,p + \frac{c}{1-|P|} (p_{12} + |P|\, p) \\ & \quad = c_1 + \frac{c}{1-|P|} (p_{12} + p) = g + h(p)\,. \end{aligned} \tag{4.19}$$

From condition (4.12) we obtain:

$$c_1 + \frac{c}{1-|P|} (p_{12} + p) \le R, \quad 0 \le p \le 1\,. \tag{4.20}$$

From (4.19) and (4.20) we have that the constant g and the function $h(p)$, $0 \le p \le 1$ satisfy the optimization equation (4.14). Also the strategy δ^∞ is AC-optimal because for each $0 \le p \le 1$ the action $\delta(p) = 0$, $0 \le p \le 1$ minimizes the second part of (4.14).

Inversely, we suppose that δ^∞ with $\delta(p) = 0$, $0 \le p \le 1$ is AC-optimal. Then in combination with optimization equation (4.14) we obtain:

$$g + h(p) = c_1 + c\,p + \sum_{\theta} \{\theta/p\}\, h(T(p,\theta)\}, \quad 0 \le p \le 1\,.$$

It is valid (see (4.19)) when:

$$g = c_1 + \frac{c}{1-|P|}\, p_{12}, \;\; h(p) = \frac{c}{1-|P|}\, p, \quad 0 \le p \le 1\,.$$

Also we have

$$c_1 + c_2 \sum_{\theta} \{\theta/p\} h(T(p,\theta)\} = c_1 + \frac{c}{1-|P|} (p_{12} + p) \le R, \quad 0 \le p \le 1\,.$$

This is obvious from condition (4.16). □

Let now consider the strategy δ^∞ where $\delta(p) = 0$ if $p = 0$ and $\delta(p) = 1$ if $p \ne 0$. Then, for any initial state p, $0 \le p \le 1$, the process will be at $p' = 0$, $t = 1, 3, 5, \ldots$, hence the average cost to be accrued using this policy is $g = \dfrac{c_1 + R}{2}$. We prove that above policy is not (AC) optimal.

Proposition 4.9. *The strategy* δ^{∞} (*stationary*)

$$\delta(p)=\begin{cases}0 & (\textit{continue}) \quad \textit{if} \;\; p=0\\ 1 & (\textit{replace}) \quad \textit{if} \;\; 0\prec p\leq 0\end{cases}$$

is not (*AC*) *optimal.*

Proof. We prove that optimization equation (4.14) is satisfied for

$$g=\frac{c_1+R}{2},\;\; h(p)=\frac{R-c_1}{2},\quad 0\leq p\leq 1,\; h(0)=0$$

and the strategy δ^{∞} is AC-optimal if and only if

$$R=c_1\,. \tag{4.21}$$

Actually, we have

$$g+h(p)=\frac{c_1+R}{2}+\frac{R-c_1}{2}=R,\quad 0\leq p\leq 1\,,$$

$$g+h(0)=g=\frac{c_1+R}{2}\,,$$

$$\begin{aligned}c_1+c\,p+\sum_{\theta}\{\theta/p\}\,h(T(p,\theta))&\\ =c_1+c\,p+\frac{R-c_1}{2}\sum_{\theta}\{\theta/p\}&\\ =c_1+c\,p+\frac{R-c_1}{2}&\\ =\frac{c_1+R}{2}+c\,p,\quad 0\leq p\leq 1\,.&\end{aligned}$$

Notice that if $p=0$,

$$c_1+c\,0+\sum_{\theta}\{\theta/0\}\,h(T(0,\theta))=\frac{c_1+R}{2}=g+h(0).$$

Therefore the optimization equation (4.14) is valid for $p=0$ if and only if:

$$g+h(0)=\frac{c_1+R}{2}\leq R\,. \tag{4.22}$$

If $p\neq 0$ the optimization equation (4.14) is valid for $p=0$ if and only if

$$g+h(p)=R\leq c_1+c\,p+\sum_{\theta}\{\theta/p\}\,h(T(p,\theta)),\quad 0\leq p\leq 1\,.$$

Therefore,

$$R\leq\frac{c_1+R}{2}+c\,p,\quad 0\leq p\leq 1\,. \tag{4.23}$$

But (4.23) is equivalent to

$$R\leq\frac{c_1+R}{2}\,. \tag{4.24}$$

From (4.22) and (4.24) implies that the optimization equation (4.14) is satisfied if and only if:

$$\frac{c_1+R}{2}=R.$$

Therefore, $R=c_1$. Because $c_1 \prec c_2 \prec R$ the condition (4.21) is not valid and therefore the above strategy δ^∞ is not average cost optimal. □

Proposition 4.10. *If*

$$(c_2 \prec)R \prec c_1+\frac{c}{1-|P|}(p_{12}+1) \tag{4.25}$$

then the stationary strategy

$$\delta(p)=\begin{cases}0 & (\textit{continue}) \quad \textit{if} \;\; 0\le p \prec p^*\\ 1 & (\textit{replace}) \quad \textit{if} \;\; p^* \prec p \le 1\end{cases},$$

is average cost (AC) optimal, where $p^*\in(0,1)$ is a critical point.

Proof. From Proposition 4.6 implies for each $\beta\in(0,1)$ the critical point of β-optimal strategy $p_0(\beta)\in(0,1)$ if and only if

$$R \prec c_1+\frac{c}{1-\beta|P|}(\beta\, p_{12}+1).$$

Because the function $f(\beta):=\dfrac{1+\beta\, p_{12}}{1-\beta|P|}\,c$, $0\prec\beta\prec 1$ is monotone increasing and

$$\lim_{\beta\to 1^-} f(\beta)=\frac{1+p_{12}}{1-|P|}\,c,$$

from (4.25) implies that exists $0\prec\varepsilon\prec 1$ such that:

$$R \prec c_1+\frac{c}{1-\beta|P|}\,(\beta\, p_{12}+1) \quad \forall\ \beta\in(1-\varepsilon,1).$$

Therefore,

$$0\prec p_0(\beta)\prec 1 \quad \forall\ \beta\in(1-\varepsilon,1)$$

Let $\{\beta_n\}\subseteq(1-\varepsilon,1)$ be such that, $\lim\limits_{n\to\infty}\beta_n=1$. Then $0\prec p_0(\beta_n)\prec 1$, $\forall\ n=1,2,\ldots$. By the Bolzano-Weierstrass theorem there is a subsequence $\{\beta_{n_k}\}$ such that:

$$p_0(\beta_{n_k})\to p^*,\ \ k\to\infty \ \ \text{and} \ \ p^*\in[0,1].$$

If $p^*\succ 0$, then for $p\in[0,p^*)$ fixed there exists an $\tau\in N$ such that:

$$p\prec p_0(\beta_{n_k}),\quad \forall\ k\ge\tau.$$

Thus it is β_{n_k}-optimal to produce at p, for all $k\ge\tau$; hence it is (AC)-optimal to produce at p, by Proposition 3.5. Since $0\le p\le p^*$ was arbitrary it is therefore (AC)-optimal to produce for $p\in[0,p^*)$. On the other hand, if $p^*\prec 1$, it is similarly shown that it is (AC)-optimal to replace for $p\in(p^*,1]$. We claim that $0\le p^*\le 1$. We argue by contradiction: $p^*\neq 0$, $p^*\neq 1$. The fact that $p^*\neq 0$, is easily implies by Proposition 4.9. We consider now $p^*=1$. Then for each $p\in[0,1)$ the decision

$\delta(p) = 0$ is (AC)-optimal. This is not valid by Proposition 4.8 and condition (4.25). Therefore $p^* \neq 1$. Hence, $0 \prec p^* \prec 1$. □

5. Optimization Procedure

Recall that $\Delta(\pi)$ is defined in Eq. (5.1) as:

$$\Delta(\pi) = g + V(\pi/\delta) - \left[\left[\pi q^{\delta_1} + \sum_{\theta}\{\theta/\pi, \delta_1\} V[T(\pi/\theta, \delta_1)/\delta\right]\right], \quad (5.1)$$

where policy δ^∞ is being improved. If the optimal gain is defined as g^* then it can be shown that:

$$\min_{\pi} \Delta(\pi) \leq g_\delta - g^* \leq \max_{\pi} \Delta(\pi). \quad (5.2)$$

The result is analogous to the completely observable result discussed by Howard [15]. With, the bound on the distance from the optimal gain in Eq. (5.2), the policy iteration algorithm is complete.

Policy-Iteration with $\beta = 1$

Initial step: Pick an arbitrary policy, say $\delta(\pi) = \alpha$, $\forall\ \pi$.

Step 1. Choose the degree of the partition k to satisfy error requirements, and find the partition V^k. k indicates the number of times that the inverse mapping, T^{-1},will be applied to the points of discontinuity of δ in order to obtain the partition for the piecewise-linear function V. There is some finite number k' after which this algorithm cannot create new boundaries, because as we proved the policy is finitely transient.

Step 2. Construct the mapping $\nu(j, \theta)$ from V^k, where $V^k = \{V_1, V_2, \ldots V_p\}$. That means we take some $\pi = (1 - \pi_2, \pi_2) \in V_j$, $j = 1, 2, \ldots, p$, and finding

$$T_2(\pi_2, \theta = 2) = \frac{r_{22}[\pi_2(1 - p_{21} - p_{12}) + p_{12}]}{(r_{11} + r_{22} - 1)[\pi_2(1 - p_{21}p_{12}) + p_{12} + 1 - r_{11}]}$$

and

$$T_2(\pi_2, \theta = 1) = \frac{a_1 + \beta_1 \pi_2}{\gamma_1 + \delta_1 \pi_2}$$

$$= \frac{\pi_2 [(1 - r_{22})(1 - p_{21} - p_{12})] + p_{12}(1 - r_{22})}{\left(\begin{array}{l}\pi_2(r_{11} + r_{22} - 1)(p_{21} + p_{12} - 1) \\ + r_{11}(1 - p_{12}) + p_{12}(1 - r_{22})\end{array}\right)}.$$

For example if $\pi \in V_2$, $T_2(\pi_2, \theta = 2) \in V_4 \Rightarrow \nu(2, 2) = 4$.

Step 3. Calculate $\overline{\gamma}_\delta$ and g_δ from

$$g_\delta \underline{1} + \overline{\gamma}_\delta = \overline{P}_\delta \overline{\gamma}_\delta + \overline{q}. \quad (5.3)$$

With the construction of ν, the matrix $\overline{P}_\delta$ of finite state controller and $\overline{\gamma}_\delta$ are well defined. Thus the equation (5.3) can be solved for $\overline{\gamma}_\delta$ and g_δ by fixing one value $\overline{\gamma}_\delta$ for each chain represented in $\overline{P}_\delta$.

Step 4. *Policy-improvement*. Find the policy $\delta_1(\pi)$, where $\delta_1(\pi)$ minimizes

$$\left[\pi q^{\alpha} + \sum_{\theta}\{\theta/\pi, a\}\, V[T(\pi/\theta, \delta)/\delta]\right] \tag{5.4}$$

over α, and where $V(\pi/\delta) = \pi\gamma_{v(\pi)}$.

Step 5. Evaluate $g_\delta - g^*$ from $\Delta(\pi)$, where

$$\Delta(\pi) = [g_\delta + \pi\gamma_{v(\pi)}] - \left[\left[\pi q^{\delta_1(\pi)} + \sum_{\theta}\{\theta/\pi, \delta_1\}\, V[T(\pi/\theta, \delta_1)/\delta\right]\right] \tag{5.5}$$

and

$$\min_{\pi}\Delta(\pi) \le g_\delta - g^* \le \max_{\pi}\Delta A(\pi)\,. \tag{5.6}$$

Step 6. If $|g - g^*| \prec \varepsilon$ then stop; the optimal policy (within ε) is δ, otherwise, return to step 1 with δ replaced by δ_1.

The mechanics of this step are essentially the same for both the discounted and undiscounted problems. The max and min of $\Delta(\pi)$ are the upper and lower bounds of $|g - g^*|$. Regardless of whether $V(\pi/\delta)$ or $\overline{V}(\pi/\delta) = \min_j[\pi\gamma_j]$ is used to calculate $\Delta(\pi)$ in *policy-improvement*, $\Delta(\pi)$ will be piecewise linear (but not necessarily continuous). Thus the maximum and minimum can occurs only at the breakpoints. By noting this fact the bounds can be readily determined as $\Delta(\pi)$ is computed.

6. **Conclusions**

This paper has discussed an optimal replacement problem of a discrete-time Markovian deterioration system with an incomplete monitoring mechanism.The objectives of this research are:

(a) to develop sufficient conditions which a control-limit policy,
(b) to investigate the structural properties for the two-state POMDP.

In this paper, it was assumed that the transition probability of the deteriorating process of the system and the probabilistic relation between the system and the monitoring mechanism are completely known.

References

[1] A. Arapostathis, Fernandez-Gaucherad and S. Markus, *Siam J. Control & Optimization* **31**(2) (1993), 282–344.
[2] K. Astrom, Optimal control of Markov processes with incomplete state information, *Journal of Mathematical Analysis and Applications* **10** (1965), 174–205.
[3] D. Blackwell, Discounted dynamic programming, *Ann. Math. Stat.* **36** (1965), 226–235.
[4] D. Bertsekas, *Dynamic-Programming*, Prentice Hall, Englewood Cliffs, New Jersey, 1997.

[5] R. Cavazos, Necessary and sufficient conditions for a bounded solution to the optimality equation in average reward Markov decisions chains, *Syst. Control Letters* **10** (1988), 71–78.
[6] C. Derman, *Optimal Replacement Rules When Changes of States Are Markovian*, *Mathematical Optimization Techniques*, University of California Press, Berkley, CA, 1963.
[7] J. Goulionis, Periodic policies for partially observable Markov decision processes, *Working paper 30* (2004), 15–28, University of Piraeus
[8] J. Goulionis, A replacement policy under Markovian deterioration, *Mathematical Inspection* **63** (2006), 46–70.
[9] J. Goulionis, POMDPs with uniformly distributed signal processes, *Spoydai* **55** (2005), 34–55.
[10] A. Maitra, Discounted dynamic programming on compact metric spaces, *Sankyā*, Ser. A **30** (1968), 211–216.
[11] G. E. Monahan, A survey of partially observable Markov decision processes, *Management Science* **28** (1982), 1–16.
[12] S. Ross, The optimal control of partially observable Markov processes over the infinite horizon, *Operation Research* **26** (1978), 282–304.
[13] S. Ross, Quality control under Markovian deterioration, *Mngmt Sci.* **17** (1971), 587–596.
[14] Y. Sawaragi and Yoshikawa, Discrete time Markovian decision processes with incomplete state observations, *Ann. Math. Stat.* **41** (1970), 78–86.
[15] J. Edward Sondik, The optimal control of partially observable Markov decision processes over the infinite horizon: discounted costs, *Operation Research* **26** (1978), 282–304.
[16] E. J. Sondik, *The Optimal Control of Partially Observable Markov Processes*, Ph.D. Dissertation, 1971, Stanford University, CA.
[17] C. White, Optimal inspection and repair of a production process subject to deterioration, *Journal of Operation Research Society* **29** (1978), 235–243.

John E. Goulionis, *Department of Statistics and Insurance Science, The University of Piraeus, 80 Karaoli & Dimitriou Street, 18534 Piraeus, Greece*.
E-mail: jgouli@unipi.gr

Dimitrios I. Stengos, *Department of Statistics and Insurance Science, The University of Piraeus, 80 Karaoli & Dimitriou Street, 18534 Piraeus, Greece*.
E-mail: sondi@otenet.gr

Received September 15, 2009
Accepted November 17, 2010