MODELING DECISION MAKING IN CLINICAL PRACTICE:A COST-EFFECTIVENESS APPROACH

John E. Goulionis[1] and A.Vozikis[2]
Department of Statistics and Insurance science
University of Piraeus, 80 Karaoli and Dimitriou Street,18534 Piraeus, Greece

ABSTRACT
Current medical research, focused on understanding the disease from a molecular level is exploring the correlation between various inflammatory markers (cytokines) and patient survival. Partially observable Markov decision processes (POMDPs) have recently been suggested as a suitable Model to formalizing the planning of Clinical Management over a prolonged period of time. In this paper, we show how the POMDP framework can be used to model and solve the problem of the management of patients, characterized by hidden disease states, investigative and treatment procedures. This model is significant because it provides a way to make a tradeoff between choosing the investigative actions and the diagnosis actions. The results in this paper demonstrate the potential value of inexpensive, accurate testing procedures as well as accurate interpretation of test results. The reported experiments show that (POMDPs) provide Clinically Reasonable and justifiable solutions.

MSC: 90B50, 65K05, 65K10, 90C39

KEYWORDS: Health economics, Partially Observable Markov Decision Processes, Stochastic Dynamic Programs, Bayesian Statistics.

RESUMEN
Las usuales investigaciones médicas, se enfocan en la comprensión de la enfermedad desde un nivel molecular al explorar la correlación entre varios marcadores inflamatorios (cytokines) y la sobrevivencia de los pacientes. Procesos de decisión de Markov parcialmente observables (POMDPs) han sido sugeridos recientemente como un adecuado modelo para formalizar el planeamiento del manejo clínico sobre un periodo prolongado. En este trabajo , demostramos como el marco de trabajo brindado por POMDP puede ser usado para modelar y resolver el problema del manejo de los pacientes, caracterizado por los estados latentes de la enfermedad, procedimientos investigativos y de tratamientos. Este modelo es significante porque provee una vía para establecer un compromiso entre seleccionar acciones investigativas y acciones diagnosticas. Los resultados de este trabajo demuestran el valor potencial de baratos así como procedimientos de prueba de interpretación acurada. Los reportes experimentales muestran que (POMDPs) provee soluciones Clínicamente Razonables y soluciones justificables.

1. INTRODUCTION

The diagnosis of a disease and its treatment are not separate processes. Although the correct diagnosis helps to narrow the appropriate treatment choices, it's often the case that the treatment must be pursued without knowing the underlying patient state with certainty. The reason for this is that the diagnostic process is not a one shot activity and it is usually necessary to collect additional information about the underlying disease, which in turn may delay the treatment and make the patients' outcome worse. This process may be even more complex when uncertainty associated with the reaction of a patient to different treatment choices and costs associated with various actions need to be considered. Thus, in a course of patient management one needs to carefully evaluate the benefit of possible diagnostic and treatment steps. To model accurately the complex sequential decision process that combines diagnostic and treatment steps, we need a framework that is expressive enough to capture all relevant features of the problem.

[1] Corresponding Author John E. Goulionis: mailto goulionis25@gmail.com tel. 00302109735066
[2] A Vozikis mailto avozik@unipi.gr tel 00302104142272

The tools typically used to model and analyze decision processes are stochastic decision trees, [14]. Unfortunately stochastic decision trees are not always the best choice, especially when a problem domain is complex and long decision sequences need to be considered. The key drawback of stochastic trees is that they require a large number of parameters to be defined, and thus, are hard to construct and modify.

At the heart of decision theoretical planning is a Model called Markov decision process. A Markov decision process (MDP), see [2,3,8,9,10] extends the classic planning framework in several aspects. First, it allows effects of actions to be nondeterministic . Second the feedback from the world provides the exact information about the states of the world. A Markov decision process describes a stochastic control processes with the assumption of perfect Observability.

Unfortunately, this assumption is too strong for many practical planning problems. Essentially, it corresponds to the situation in which we know with certainty what is the disease (or complication) the patient suffers from at any point in time. Also, if the disease is always known with certainty, there would be no need for investigative actions or procedures. This is in contrast to many medical problems in which investigative procedures are very common and play a major role.

A framework more suitable for modeling the outlined therapy problem is partially observable Markov decision process (POMDP), see [2,3,4,5,6,7]. A POMDP represents a controlled Markov process with two sources of uncertainty. Stochastic related to the dynamics of the control process, and uncertainty associated with the partial observability of the disease process by a decision maker. The model of POMDPs is a generalization of MDPs. It assumes that the effects of the actions are nondeterministic as in MDP but does not assume that feedback provides perfect information about the state of the world. Instead, it assumes that feedback may provide incomplete or imperfect information about states of the environment, see c.f [11].

In this paper we apply POMDPs to medical therapy planning for patients. To reduce the computational complexity of problem–solving methods, we applied completely unobservable state. To build and solve the problem, we use a factored version of a POMDP with a hierarchical dynamic belief network model of the disease dynamics. Using structure–based techniques we were able to construct a model of moderate complexity and successfully solve a number of cases. To see why this paper is important, let us take a look at the interactions between an agent modeled by POMDP and its environments.

On one hand, the patients' state can be changed by executing actions. This procedure can be viewed as the control effects of actions. On the other hand, a feedback is provided to the agent when the patients' states change. This procedure can be viewed as the information-gathering effects of actions since different actions can change patients to different states and in turn this allows the agent to receive different feedback. Therefore we provide a unified framework to handle these two sources of uncertainties: the control effects of actions and information gathering effects of actions. Hence, the model provides one way to make a tradeoff between choosing actions to change the patients' states and actions to collect information for the agent. This tradeoff will be highlighted by one example in next section.

The reminder of this paper is organized as follows. Section 2 formally describes the Markov Model and the first steps in building a general model. In section 3 we give the expressiveness of the POMDP framework and we give structural results describes model results with respect to changes in test cost and accuracy. Due to the finite observation and treatment horizons, the model is formulated as a finite horizon POMDP, where the patient's true health state can only be observed through an inaccurate testing procedure that measures the value of a single cytokine level. In section 4 we present an algorithm for solution our problem and finally in section 5, for illustration a simple numerical example

2. MEDICAL DECISION MAKING

The first step in building a Markov Model of prognosis is the enumeration off all distinct states of health. For example Well, Ill and Dead. Importantly, they should enable estimation or abstraction from the literature of specific transition probabilities per unit time among the various states. Some transitions in the Markov model may reflect two or more independent forces. We may suppose that the transition from Well to Dead reflects both the force of mortality in the general population and disease-specific

effects. Methods for assessing disease specific mortality rates have been reported elsewhere, see Anderson [1].

For Markov medical decision making we have found it useful to consider two types of states: long term and temporary. Long term states, are states in which it is possible to remain from cycle to cycle. Temporary states reflect short-term events that force transition to another state in the model in the next cycle. For example, we could add a state to the model labeled HOSPITALIZED, to reflect one cycle spent in the hospital .If all transitions from HOSPITALIZED were to other states, then this new addition would be a temporary state.

Note that in our temporary state the probability of remaining in that state is set to zero. For example, three such states might be HOSPITALIZATION, Period 1, HOSPITALIZATION, Period 2 and HOSPITALIZATION, Period 3.

Life expectancy, a common employed outcome measure for clinical Decision making is defined as the average future lifetime of a cohort of patients with identical clinical features. A large number of patients are followed as a cohort. The cohort begins in an initial distribution of states, and at each cycle of the process the entire cohort is reallocated to states according to the transition probabilities. In the cohort analysis two values are kept for each state to each cycle: the number of patients currently in the state and the total number of patient-cycles in the state. For example, at cycle 2 the Well state includes 9000 patients; the cumulative number of Well patients months is 100000+30000+9000, or 139000.

After a sufficient number of cycles almost everyone will be dead. The few remaining patients can be treated in one of two ways: a small arbitrary amount of life expectancy can be added to each state that has remaining cohort members or the life expectancy can be truncated. The Markov cohort simulation is stopped when the remaining cohort has diminished to the point where any error introduced by summing up the experience of the remainder (dealing with the tail) is small compared to the total patient cycles accumulated during the analysis.

When a Markov cohort analysis is terminated, the total number of patient-cycles for each state is divided by the size of the original cohort, yielding the expected time that each individual member will spend in each state. Life expectancy is the sum of these expected values.

Table 1. Markov Cohort simulation

Time	State WELL	 ILL	 DEAD
0	100000	0	0
1	30000	50000	20000
2	9000	40000	51000
3	2700	24500	72800
.	.	.	.
.	.	.	.
.	.	.	.
Sum Average Cycles (Sum/100000) Life expectancy	142860 1.43 1.43+1.43=2.86 cycles	142860 1.43	....

2. MODEL DESCRIPTION AND ASSUMPTIONS

A model that remedies the disadvantages of perfectly observable MDPs and still preserves some of their good features, like time-decomposability and reduced model complexity is POMDP see [2, 3,11,12]. The POMDP generalizes the MDP framework by allowing the patient Health state to be partially observable through a testing process [13,15,19]. A Partially Observable Markov Decision Process, POMDP , is a collection *(S, D, P, Θ, R, C, β).* The POMDP consists of a core process x_t , an observation process z_t , and a decision process d_t .

The core process $x_t, t = 0,1,2,\ldots$ is a discrete–time Markov process.

- S={0, 1, 2, 3, …, N} is a finite state of process states (disease states); where 0, is the good state and N is the Worst state. At any given time period, the decision maker selects one of the actions of the set D.

- D is the action space, (diagnostic and treatment procedures); and Θ is a finite set of observations (findings, results of diagnostic tests); All are finite and the stage invariant. D is taken to be composed of three disjoint sets D_{test}, D_{treat}, D_{skip}, where D_{test} constitutes the set of available diagnostic procedures, D_{treat} ={Surgery} and D_{skip} is a singleton set that consists of the special action skip (i.e. refrain from acting at the specified point in time) only. For example,Ischemic Heart Disease is caused by an imbalance between the supply and demand of oxygen to the heart. The condition is most often caused by the narrowing of coronary arteries (coronary artery disease) and an associated reduction in the oxygenated blood flow. The coronary artery disease tends to progress over time. The pace of the disease progress is stochastic and contingent on multiple factors. We consider at any point in time, the physician has different options to intervene:

D_{treat}={angioplasty–PTCA, coronary artery bypass surgery–CABG}

D_{test}={stress test, coronary angiogram}, that tends to reveal more about the underlying status of the coronary disease. Some of the interventions have a low cost, but some carry a significant cost associated with the invasiveness of the procedure.

D_{skip}={medication treatment, do nothing}.

The state of the patient undergoes deterioration according to a stationary discrete-time Markov chain having a known transition law. Let p_{ij} denote the1-step transition probability from state i to state j.

- p_{ij}^{d} =pr[x_{t+1}=j /x_t= i, d_t=d], P^d=(p_{ij}^{d})

- 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 Θ={1,2, …,M}. It's assumed that the probabilistic relation between the state of the patient and the outcome of the monitoring is prescribed by the following known conditional probability:

$r_{i\theta} = \Pr$ the outcome of the monitoring is level θ | the patient is in state i ,

$i = 1,2$, …, N. $\theta = 1,2,\ldots,M$.

r_{iq}^{d} =Pr[$z_{t=}\theta$/ x_t= i, d_{t-1}=d] $i \in S$,$d \in D$.

The basic POMDP model includes an observation process that relates observable information (i.e. results) to the patient's true cytokine levels through the above known probability distribution.

- R^d is the observation matrix. The observation is the outcome of test or response to therapy, where $r_{j\theta}^{d}$ =Pr[Y_t=θ/X_t=j, d_{t-1} =d], $j \in$ S, $d \in$ D is the probability that observation $\theta \in \Theta$ ={1,2,3…,M} will be observed at the next stage, if the state at the next stage is j and d is the current action.
- β is a discount–factor $0<\beta \leq 1$ for finite horizon c(i, d), $0 \leq t<K$ is the scalar –valued cost accrued, when the current state is. $i \in$ S, and the action $d \in$ D. The cost for a transition from state s to state s´ under action d consists of three components:

C(s, d, s´)**:** =C (s´)+C (d) , (1)

where *C(s´)* is a cost associated with a patient state only, *C(d)* stands for a cost associated with an action (e.g. cost of a surgery), that includes the economic cost and patient's discomfort. The cost, *C(s´)*, can be decomposed to costs associated with the amount of chest pain the patient suffers at a given time, and a loss $l_\kappa(s')$ to denote life expectancy (in years) associated with final state s´ at time $t=K$, where no action choice is made.

$$C(s') = \sum_i [C(s_i') + l_\kappa(s_i')] \quad (2)$$

For initial information vector (belief–state) we note that the current prior can be found by a more flexible model that targets different groups of patients and exploits other context information. For example sex, age, smoking history, etc. There are logistic regression models developed for this purpose, see [12]. The cost structure considered here is as follows: $C^d(i)$, where $c(i,d)$ is the scalar valued cost accrued, when the current state is $i \in S$ and action is $d \in D$. Because a test result may not received in every stage and the tests results that are received may either be inaccurate themselves or interpreted inaccurately, the model uses a belief vector to describe a probability distribution over the possible core states.

An observer does not directly observe the core process. He sees instead one of the outputs which is probabilistically related to z_t. Although the state of the core process is not known with certainty, it is possible to calculate the probability that the core process is in a given state. In particular we define**:**
$\pi_i(t) = \Pr(x_t = i \mid z_0, \ldots, z_t, d_0, \ldots, d_{t-1})$

The vector $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_N(t))$ is called information vector, and the space of all such vectors, Π, is called the information space.

We have: $\sum_{i=1}^{N} \pi_i(t) = 1$ and $\pi_i \geq 0$. It is well known that $\pi(t)$ is a sufficient Statistic [16,17].

More precisely, $\pi(t)$ summarizes all of the necessary information of the history of the process for choosing an action at time t.

The values of the vector correspond to the clinician's belief that the patient's true health is in each of the possible states. The model utilizes an observation probability matrix to relate the observed values to the underlying health state. Then, using a prior estimate of the belief vector the current tests results, and knowledge of the last action taken, Bayesian updating is used to form a new estimate of the belief vector. The clinician's decision is made based on the value of this belief vector at each decision point. If the information vector at time t is π and an alternative α is selected, and if an output θ results, then the new information $\pi(t+1)$ is given by $T(\pi \mid \theta, d)$.

By Bayes' rule.

$$T(\pi(t) \mid \theta, d) = \pi(t+1) = \frac{\pi \cdot P^d \cdot R_\theta^d}{\{\theta \mid \pi, d\}}. \quad (3)$$

$\theta \mid \pi, d = \pi \cdot P^d \cdot R_\theta^d \cdot \underline{1}$ is the probability of receiving observation θ at stage $t+1$, given that $\pi(t)$ and d is the action selected at stage t.

R_θ^d be the diagonal matrix having $r_{i\theta}^d$ as its j-th diagonal term and zeros for all off– diagonal terms.

Assuming $\underline{1} = col(1, \ldots, 1)$.

The objective of a POMDP is to find an optimal policy among the admissible policies such that it minimizes a given performance index, typically the total expected discounted cost to be accrued over the infinite horizon, or the expected long-run average cost, conditioned on the a priori $\pi(0)$. These

costs are defined in terms of the state x_t for each admissible strategy, δ, and information vector π (0) of the initial state by:

The total expected discounted value for an initial information vector $\pi \in \Pi$ can be stated as:

$$V^*(\pi)=\min_{(\delta_0,\delta_1,\ldots)} E\left[\sum_{t=0}^{K} \beta^t . \pi(t) . C_{\delta_t(\pi(t))} / \pi(0)=\pi\right], \quad (4)$$

where the sequence of control function $\{\delta_0, \delta_1, \ldots\}$ is termed a policy.

The objective of the therapy planning (finite horizon) is to develop a strategy that would minimize the expected cumulative cost of the treatment, where the cost is defined in terms of the dead-alive trade-off, quality of life, invasiveness of procedures and their economic cost.

4. SOLUTION METHOD

The standard approach to solving POMDP problems was initiated by Sondik [18]. He showed that the minimum expected of operating the POMDP for a finite time period is a piecewise-linear concave function over Π. Let V_0^* be any arbitrary bounded function, then for the finite horizon case

$V_{t+1}^*(\pi) = \min\{\pi.\gamma: \gamma \in \Gamma_t\}$ for some finite set Γ_t of vectors in R^n.

Using this representation for V_t in the Dynamic programming (D.P) recursion

$$V_{t+1}^*(\pi)=\min\{\pi \cdot [c^d+\beta \cdot P^d . \sum_{\theta=1}^{M} R_\theta^d \cdot \gamma^{l(\pi,d,\theta)}]: d \in D\} \quad (5)$$

where $l(\pi,d,\theta)$ is the index of the $\gamma \in \Gamma_t$ that minimizes $\pi \cdot P^d \cdot R_\theta^d \cdot \gamma$, see [16],[15].

Thus, given Γ_t and any $\pi \in \Pi(s)$, one would find $l(\pi,d,\theta)$ for each $d \in D$ and $\theta \in \Theta$, and then find the optimal action and $V_{t+1}(\pi)$ from relation(5).

The algorithm to discover all of the vectors in the set Γ_{t+1} appropriate for defining V_{t+1}^* proceeds by choosing an arbitrary starting mass function π_0 in the unit simplex, finding the indices $l(\pi_0, d, \theta)$ for each d and θ, and then finding the optimal action d^* in D from equation (13). Then the vector in the inner brackets in (5) evaluated at d^* and π_0, call it γ^*, satisfies

$$V_{t+1}^*(\pi_0) = \gamma^* . \pi_0. \quad (6)$$

This new vector γ^* is added to the Γ_{t+1} set. The boundaries of the region in R^n over which γ^* is operative as the gradient of V_{t+1}^* are discerned by considering how the inner bracketed quantity in (5) might change as the mass vector π changes from the initial π_0. Two ways are suggested: either the index $l(\pi,d^*,\theta)$ changes from $l(\pi_0,d^*,\theta)$, or the optimal action changes. This first possibility will be realized if

$$\pi \cdot P^{d^*} \cdot R_\theta^{d^*} \cdot \gamma^{l(\pi_0,d^*,\theta)} < \pi \cdot P^{d^*} \cdot R_\theta^{d^*} \cdot \gamma \quad \text{for } \gamma \in \Gamma_t$$

consequently the set of linear inequalities:

$$\pi \cdot [P^{d^*} \cdot R_\theta^{d^*} \cdot \gamma^{l(\pi_0,d^*,\theta)} - P^{d^*} \cdot R_\theta^{d^*} \cdot \gamma] \le 0 \quad \text{for } \gamma \in \Gamma_t, \theta = 1, 2, \ldots M \quad (7)$$

defines the region over which this possibility is excluded. The second possibility is investigated in the algorithm by setting up the inner bracketed expression in (5) for $\pi=\pi_0$ and each action *d*, call it $\gamma(d)$, and expressing the region over which d^* is operative by the set of linear inequalities:

$\pi[\gamma^* - \gamma(d)] \le 0$, with $d \in D$, (8)

where $\gamma(d) = [C^d + \beta \sum_{\theta=1}^{M} p^d \cdot R_\theta^d \cdot \gamma^{l(\pi_0,d,\theta)}]$.

Note that with this terminology $\gamma^* = \gamma(d^*)$. The complete set of inequalities (7) and (8) are supposed to characterize the region over which γ^* is optimal and γ^* is the gradient of V_{t+1}^*. This note now diverges from the Sondik paper, see [17]. The set of inequalities (7) and (8) does not necessarily capture the intended region, because in constructing $\gamma(d)$, $l(\pi 0,d,\theta)$ is used and it's possible that for some π satisfying the inequalities (8) we might have $l(\pi,d,\theta) \neq l(\pi 0,d,\theta)$ and the inner bracketed term in (5) evaluated at action d and $l(\pi,d,\theta)$, call it $\gamma''(d)$,can satisfy

$$\pi \cdot \gamma''(d) < \pi \cdot \gamma^* \leq \pi \cdot \gamma(d) \tag{9}$$

for some $d \in D$. There will then exist a $\gamma''(d)$ vector that should be added to Γ_t as operative at π, but since the inequalities (8) will be satisfied the boundary will be missed. This possibility can be prevented by allowing for all possible combinations of action and $l(\pi,d,\theta)$ indices in the inner bracketed term in (8),i.e. all possible components of Γ_t. Letting $\#D$ denote the cardinality of the set D, there will be $(\#D) \cdot (\#\Gamma_t^M)$ possibilities. Let G denote this set of all possible gradient vectors for period $t+1$. Then starting with π_0 and calculating γ^* as above, we have that γ^* is the operative slope if and only if

$$(\gamma^* - \gamma) \cdot \pi \leq 0, \quad \gamma \in G. \tag{10}$$

This extended set of inequalities will detect all possible changes, and hence subsumes all of the inequalities (7), (8) and (10).

5. NUMERICAL EXAMPLE

In order to illustrate the solution procedure for the problem we present a numerical example. The parameters of problem are given as follows.
$S=\{1,2,3\}$, $\Theta=\{1,2,3\}$, $D=\{0,1\}$, discount factor $\beta = 0.95$.

The transition matrices are the following for d=0 and d=1 respectively: $P^0 = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.05 & 0.75 & 0.2 \\ 0 & 0.15 & 0.85 \end{pmatrix}$

, $P^1 = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.7 & 0.2 & 0.1 \\ 0.4 & 0.6 & 0 \end{pmatrix}$

The observation matrices are the following for α=0 and α=1 respectively:

$$R^0 = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1. & 0.8 \end{pmatrix}, R^1 = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

Hence,

$$R_1^0 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & .0 & 0.1 \end{pmatrix}, R_2^0 = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & .0 & 0.1 \end{pmatrix}, R_3^0 = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & .0 & 0.8 \end{pmatrix}$$

$$R_1^1 = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & .0 & 0.1 \end{pmatrix}, R_2^1 = \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & .0 & 0.2 \end{pmatrix}, R_3^1 = \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & .0 & 0.7 \end{pmatrix}$$

The initial information vector is $\pi(0) = (0.3, 0.5, 0.2)$.

$$c^0 = \begin{pmatrix} 1000 \\ 3000 \\ 7000 \end{pmatrix},\ c^1 = \begin{pmatrix} 7000 \\ 8000 \\ 10000 \end{pmatrix}$$

In this portrayal, the space of possible π vectors is represented by an equilateral triangles, with each point in the triangle corresponding to a possible state of the information vector π. For each information vector $\pi = (\pi_1, \pi_2, \pi_3)$, the perpendicular distance from the point to the side opposite the i-th vertex is just equal to π_i ($i = 1, 2, 3$). Thus, points closer to the *i-th* vertex correspond to states of information in which the process is believed more likely to be in state *i*.

Optimal control-limit policy: A line segment connecting information vectors
$\pi' = (0.50, 0.50, 0)$ and $\pi'' = (0.35, 0, 0.65)$

5. CONCLUSIONS

In this paper we provide a method to compute the optimal cost and policy for clinical patient management, using the model of POMDPs. This model provides an elegant framework for medical therapy planning problems, to finite horizon and it may be able to inform clinical practice by providing policy results in addition to suggesting general strategies. Consequently, our models should only be used as an aid in the decision making process. Future research efforts could include the addition of multiple variables in the state description, including completely observable and partially observable elements.

RECEIVED APRIL 2010
REVISED JULY 2011

REFERENCES

[1] ANDERSON, R.N.(2002): Deaths: Leading causes for 2000. **National Vital Statistics** Reports 50 (16) 1–86 .

[2] BERTSEKAS D.P. (1995): **Dynamic programming and Optimal Co**ntrol. Bemont, MA;Athena Scientific .
[3] BERTSEKAS, D.P. (2011): **Dynamic Programming and Optimal Control**. Volumes 1 and 2. Athena Scientific,

[4] BURNS, L., D. WHOLEY (1991): The effects of patient, hospital, and physician characteristics on length of stay and mortality. **Medical Care** 29 , 225–271.

[5]**.** GOULIONIS J.E., and STENGOS D.I. (2008) :The Partially observable Markov decision processes framework. **Medical decision Making. Advances and Applications in Statistics**, 9, 205-232 (2008).

[6]**.** GOULIONIS J.E. & A. VOZIKIS (2009): Partially Observable Markov decision Processes in the therapy of Parkinson's disease under average cost criterion. **Australian and New Zealand Health Policy**. 6:1-15.

[7] **.** GOULIONIS J.E.,V.K.BENOS, D. NIKOLAKIS and A. VOZIKIS (2010): On the decision rules of cost effectiveness treatment for patients with diabetic foot syndrome. **Clinicoeconomics and Outcome Research**. Final Acceptance.

[8]**.** GOULIONIS J.E. (2009): Medical treatments using partially observable Markov Decision Processes**. JP Journal of Biostatistics**. **3**, 77-97.

[9]**.** GOULIONIs J.E. (2005%: POMDPs with uniformly distributed signal processes. **Spoydai** 55, 34-55.

[10] SINGER D.E.(1998): Time to clinical stability in patients hospitalized with community-acquired pneumonia. **Journal of the American Medical Association** 279 ,1452–1457.

[11] **.** HAUSKRECHT M. (1997): Planning and control in stochastic domains with imperfect information **Phd thesis, Massachusetts Institute of Technology**.

[12] HU, C., W.S. LOVEJOY, S.L. SHAFER. (1996):Comparison of some suboptimal control policies in medical drug therapy. **Operations Research** 44 **,** 696–709.

.

[13] JOHNSTON, J.A.(2005): Determinants of mortality in patients with severe sepsis. **Medical Decision Making** 25 374–386.

[14]**.** LITTMAN M.L.(1996) Algorithms for sequential decision making. **Phd thesis, Brown University** .

[15] LOVEJOY W.S. (1991): A survey of algorithmic methods for partially observed Markov decision processes. **Ann. Operation Research**, 28, 47-66.

[16]**.** SONDIK.J.E. (1978): Optimal control of partially observable Markov decision processes over infinite horizon**. Qperations Research**, 26, 282-304.

[17] SONDIK J.E. (1971): Planning and control in stochastic domains with Imperfect information. **Phd thesis, Stanford University.**

[18] **.** SONDIK J.E. (1973): The optimal control of partially observable Markov decision processes over finite horizon. **Operations Research** , 21, 1071-1088.

[19] WHITE D.J.(1981): Infinite horizon Markov decision processes with unknown or, Variable discount factors. **European Journal of Qperations Research**, 28, 96-108 .