LINEAR PROGRAMMING APPROACH TO MARKOV DECISION MODEL FOR HUMAN HEALTH

Abubakar, U. Y. (Ph.D)
Department of Mathematics/Statistics,
Federal University of Technology, Minna, Nigeria.

Abstract
A linear programming method and the application to a Markov decision model of human health conditions are discussed in this paper. This approach involves the use of simplex method to solve a linear program model formulated with discrete time Markov decision processes in contrast to the policy- iteration algorithm. The model was initially formulated and solved with policy iteration method and now verified on identical data using linear programming. The result obtained is in agreement with the policy- iteration method. However, the linear programming method has the advantage of wide spread and simple computer software that can easily be used, unlike the policy- iteration algorithm that may demand writing its own computer codes by the individual. It is important to observe that the two methods are very efficient to determine the long-run fraction of time that a man is in a poor condition of health.

Keywords: Linear programming, Policy- iteration, Markov, Decision, Model, Health, Condition

Introduction

Linear programming is an optimization technique. It receives so much attention in recent years due to the availability of the methods of solution to the general linear programming problems involving large variables Diego and German (2006) and Abubakar(2005). Linear programming formulation of Markov Decision processes has been reported also in Diego and German (2006) and Tijm(1988). The application of Markov decision model to study human health conditions is discussed in Abubakar(2011). In that work, policy- iteration was used and found to be very involving and cumbersome, it is therefore necessary to seek for alternative method and that is the issue addressed in this paper; the linear programming approach.

Markov Decision Processes and Linear Programming.

According to Kurkani (1999), Puterman (1994), Goto et al (2004) and Hillier and Lieberman(1980); we consider a Discrete Time Markov Chain (DTMC) $\{X_n, n = 0, 1,\ldots\}$, whose transition probability matrix depends on the action taken A_n. Additionally, the system incurs a cost $c(i, a)$ when an action a is chosen at some state i. Then the joint process $\{(X_n, A_n), n=0, 1,\ldots\}$, is a Discrete Time Markov Decision Process (DTMDP).

The policy-iteration algorithm solves the following average cost optimality equation in a finite number of steps by generating a sequence of improving policies.

It was observed in Abubakar(2011) that the finite convergence of the policy-iteration algorithm implies that numbers g^* and v_i^*, $i \in I$, exist which satisfy the average cost optimality equation

$$v_i^* = \min_{a \in A(i)} \{c_i(a) - g^* + \sum_{j \in I} p_{ij}(a) v_j^*\}, \; i \in I \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

I is the set of states. The constant g^* is uniquely determined as the minimum average cost per unit time, that is

$$g^* = \min_R g(R)$$

Moreover, each stationary policy R^* such that the action R_i^* minimizes the right side of (1) for all $i \in I$ is average cost optimal Tijm (1988).

Another convenient way of solving the optimality equation is the application of a linear programming formulation for the average cost case.

According to Diego and German (2006), the next model specifies how to obtain an optimal average cost using linear programming tools.

Min $\sum_{i \varepsilon S} \pi_i \sum_{a \varepsilon A(i)} f(i,a) c(i,a)$

subject to

$$\pi_j = \sum_{a \varepsilon A(i)} \pi_i p_{ij}(a) \quad j \in S \text{ Balance equation}$$

$$\sum_{a \varepsilon A(i)} f(i,a) = 1$$ Normalization equation, S is the set of all allowable states.

This model is not linear. But if we define new decision variable $x_{ia} = \pi_i f(i,a), i \in S, a \in A(i)$, then we can build an equivalent linear model. The meaning of x_{ia} is the long run fraction of the time that the system is in state i and action a is chosen.

Min $\sum_{i \varepsilon S} \sum_{a \varepsilon A(i)} c(i,a) x_{ia}$

subject to

$$\sum_{a\varepsilon A(i)} x_{ja} - \sum_{i\in S^-(j)} \sum_{a\varepsilon A(i)} p_{ij}(a)\, x_{ia} = 0 \quad j \in S$$

$$\sum_{i\varepsilon S} \sum_{a\varepsilon A(i)} x_{ia} = 1$$

$$x_{ia} \geq 0, \qquad i \in S, a \in A(i)$$

Where $S^-(j)$ is the set of possible predecessors of state j.

i.e. $S^-(j) = \{i: j \in S(i,a) \; for\; some\; a \in A\}$. Once the model is solved, to recover the quantities of interest you must follow the following steps:

- Stationary Distribution

  $$\pi_i = \sum_{a\varepsilon A(i)} x_{ia}, \quad i \in S$$

- Value Function: This is the optimal objective value obtained by linear programming, note that the same value function applies for each state, due to be solved for the average problem.
- Decision Rule: It can be shown that there exists a deterministic decision rule, instead of a randomized one. If the transition probability matrix of every stationary policy is irreducible, the next statement shows how to get in general way.

  $$f(i,a) = \frac{x_{ia}}{\pi_i}$$

  However, if there is no knowledge about how is the performance of the discrete time Markov decision process (DTMDP), the next statement could be used, due to a DTMDP always obtain a deterministic decision rule.

$$d^*(i) = \begin{cases} a & if\; x_{ia} > 0, i \in S^* \\ arbitrary & if\; i \in S - S^* \end{cases}$$

Where,

$$S^* = \{s \in S : x^*(i,a) > o \; for\; some\; a \in A(i)\}$$

Denardo and Fox(1968) gives the following linear programming algorithm which was used in this work

Linear programming algorithm

Step 1: Apply the simplex method to compute an optimal basic solution (x^*_{ia}) to the linear program

Minimize $\sum_{i\varepsilon S}\sum_{a\varepsilon A(i)} c(a)x_{ia}$ .... (2)

subject to

$$\sum_{a\varepsilon A(j)} x_{ja} - \sum_{i\varepsilon S}\sum_{a\varepsilon A(i)} p_{ij}(a)x_{ia} = 0, \quad j \in I$$

$$\sum_{i\varepsilon S}\sum_{a\varepsilon A(i)} x_{ia} = 1,\ x_{ia} \geq 0, \quad i \in I \text{ and } a\varepsilon A(i)$$

Step 2: Start with the non-empty set

$$S: = \left\{ i / \sum_{a\varepsilon A(i)} x^*_{ia} > 0 \right\}$$

and, for any state $i \in S$, set the decision

$$R^*_i := a \text{ for some } a \text{ such that } x^*_{ia} > 0$$

Step 3: If S = *I*, then the algorithm is stopped with the average cost optimal R*. Otherwise, determine some state $i \in S$ and action $a\varepsilon A(i)$ such that $p_{ij}(a) > 0$ for some j $\in S$, set $R^*_i :=$ a and $S := S \cup \{i\}$, and repeat *step 3*.

The object of the linear program is the minimization of the long-run average cost per unit time, while the first set of constraints represent the balance equations requiring that for any state $j \in I$ the long-run average number of transitions from state j per unit time must be equal to the long-run average number of transitions into state j per unit time. The last constraint obviously requires that the sum of the fraction x_{ia} must be equal to 1.

Next we sketch a proof that the above linear programming algorithm leads to an average cost optimal policy . Following Tjims(1988), the starting point is the average cost optimality equation (1)

Since this equation is solvable then the linear inequalities

$$g + v_i - \sum_{j\in I} p_{ij}(a)v_j \leq c_i(a) \qquad i \in I \text{ and } a\varepsilon A(i) \text{ .... (3)}$$

must have a solution. Next it readily verified that any solution $\{g, v_i\}$ to this inequalities satisfies $g \leq g_i(R)$ for any $i \in I$ and any policy R, where $g_i(R)$ denotes the long-run average cost per unit time under policy R when the initial state is i. The inequalities $g \leq g_i(R)$ follow by a repeated application of the inequalities $g + v_i - \sum_{j\in I} p_{ij}(R_i)v_j \leq c_i(R_i), i \in I$; Hence we can conclude that for any solution $\{g, v_i\}$ to the linear inequalities (3) holds that $g \leq g^*$ with

g^* being the minimal average cost per unit time. Hence, using the fact that relative values v_i^* exist such that $\{g^*, v_i^*\}$ constitutes a solution to (3), linear program.

Maximize g .... (4)

subject to

$$g + v_i - \sum_{j \in I} p_{ij}(a) v_j \leq c_i(a) \qquad i \in I \text{ and } a \varepsilon A(i), g, v_i \text{ unrestricted}$$

has the minimal average cost g^* as the optimal objective-function value. Next observe that the linear program (2) is the dual of the primal linear program (4) by the dual theorem of linear programming, the primal and the dual linear program have the same optimal objective function value. Hence the minimal objective function value of the linear program (2) yields the minimal average cost g^*. To show that an optimal basic solution (x_{ia}^*) to the linear program induces an average cost optimal policy, we first prove that the non empty set

$$S_0 = \left\{ i / \sum_{a \varepsilon A(i)} x_{ia}^* > 0 \right\}$$

is closed under any stationary policy. The proof proceeds by contradiction. Suppose that $p_{ij}(a) > 0$ for some $i \in S_0$ and $j \notin S_0$ then it follows from the constraint of the linear program (3) that $\sum_{a \varepsilon A(i)} x_{ja}^* > 0$, contradicting $j \notin S_0$. By the closeness of the set S_0 under any policy and the assumption that every average cost optimal policy has no two disjoint closed sets. The states $i \notin S_0$ are transient under every average cost optimal policy. This result guarantees that the completion of policy R^* in steps 3 of the linear programming algorithm is feasible. It remains to prove that the constructed policy R^* is average cost optimal. To do so, let $\{g^*, v_i^*\}$ be the particular optimal basic solution to the primal linear program (4) such that this basic solution is complementary to the optimal basic solution (x_{ia}^*) of the dual linear program (2) then, by the complementary slackness property of linear programming

$$g^* + v_i^* - \sum_{j \in I} p_{ij}(R_i^*) v_j^* = c_i(R_i^*) \qquad \text{for} \quad i \in S_0$$

By the construction of policy R^* and the fact that the set S_0 is closed under any policy, we have that the set I(R^*) of recurrent state of policy R^* is contained in the set S_0. Thus, noting that no transition is possible from a recurrent state to a transient state.

$$g^* + v_i^* - \sum_{j \in I(R^*)} p_{ij}(R_i^*) v_i^* = c_i(R_i^*) \qquad \text{for} \quad i \in I(R^*)$$

By iterating these equalities, we find that under policy $R^*$the average cost per unit times equals g^* for each recurrent initial state. Hence, since for any transient initial state the close set of recurrent states will be reached after finitely many transitions, the average cost per unit time under policy R^* is equal to g^* for each initial state, and so policy R^* is average cost optimal.

The Model

According to Abubakar(2011), suppose that at the beginning of each day the health condition of a man is observed and classified as good health or poor health. If he is found to have poor health, he is given either a first aid/preventive treatment or curative treatment so that the health condition could change to good health and could attend to his usual activities .

Suppose also that he could be found in good health conditions i = 1,2,... N. The good health condition i is better than i+1. That is, the health condition deteriorates in time. If the present condition is i and does not fall ill, then at the beginning of the next day then he has good health conditions j with probability p_{ij}. It is assumed that his body cannot improve on its own. That is $p_{ij} = 0$ for j<i so that $\sum p_{ij} = 1$ for j>i. Let the health condition i = N represents a poor condition that requires treatment taking two days. For the intermediate states i with 1<i<N there is a choice for him to preventively take treatment so that he could remain in good health condition for the present day. Let a first aid/preventive treatment takes only one day at most and a change from poor health to a good health (after treatment) has a good health condition i=1. We wish to determine a rule which minimizes the long-term fraction of time the man is taking treatment.

Let us put the problem in the frame work of a discrete-time Markov decision model. We assume a cost of one for each day he takes treatment, the long-term average cost per day represent the long-term fraction of days that he takes treatment. Also, since a treatment for poor health condition N takes two days and in the discrete Markov decision model the state of the system has to be defined at the beginning of each day. We need auxiliary state for the situation in which a treatment is in progress. Thus the set of possible states of his health condition is chosen as

I = {1, 2, ... N, N+1}. Here the state i with 1 ≤ i ≤ N corresponds to the situation in which an observation reveals good health condition i, while state N+1 corresponds to the situation in which treatment is in progress already for one day. Denoting the two possible actions by

$$a = \begin{cases} 1 & \textit{if the condition is good health.} \\ 0 & \textit{otherwise} \end{cases}$$

The set of possible actions in state i is chosen as

A (1) = {0}, A(i) ={0,1} for 1<i<N, A(N) = A(N+1) = {1}

We find that the one step transition probabilities P_{ij} (a) are given by

P_{ij} (1) =1 for 1<i<N

$P_{N,\ N+1}(1) = 1 = P_{N+1,\ 1}(1)$

$P_{ij}(0) = P_{ij}$ for 1≤i≤N and j≥i

P_{ij} (a) = 0 otherwise

Further, the one step costs C_i (a) are given by

$C_i(1) = 1$ and $C_i(0) = 0$.

A rule or policy for controlling the health condition is a prescription for taking actions at each decision epoch.

In view of Markov assumption, and the fact that the planning horizon is infinitely long, we shall therefore consider stationary policies. A stationary policy R is a rule that always prescribes a single action R_i whenever the system is found in state i at a decision epoch.

The rule prescribing a treatment or poor health condition only when he has a good health condition for at least 5 working days is given by R_i = 0 for 1≤i<5 and R_i = 1 for 5 ≤ i ≤ N+1 .

Illustration

The average cost optimal when the number of possible working conditions equals N = 5 and the deterioration probabilities of the health conditions of staff in a company is given below

$$P = \begin{pmatrix} 0.80 & 0.15 & 0.05 & 0 & 0 \\ 0 & 0.60 & 0.20 & 0.10 & 0.10 \\ 0 & 0 & 0.40 & 0.35 & 0.25 \\ 0 & 0 & 0 & 0.50 & 0.50 \end{pmatrix}$$

The policy – iteration algorithm is initialized with the policy which prescribes treatment, be it a first aid or curative action a=1 in each state except state 1

The linear programming problem is

Minimize $\sum_{i=2}^{N+1} x_{i1}$

subject to

$$x_{10} = \left(p_{11}x_{10} + \sum_{i=2}^{N-1} x_{i1} + x_{N+1.1} \right) = 0$$

$$x_{j0} + x_{j1} \sum_{i=1}^{j} p_{ij}x_{i0} = 0 \qquad 2 \leq j \leq N-1,$$

$$x_{N1} - \sum_{i=1}^{N-1} p_{iN}x_{i0} = 0,$$

$$x_{N+1.1} - x_{N1} = 0,$$

$$x_{10} + \sum_{i=2}^{N-1} (x_{i0} + x_{i1}) + x_{N1} + x_{N+1.1} = 1,$$

$$x_{10}, x_{i0}, x_{i1}, x_{N1}, x_{N+1.1} \geq 0.$$

The linear program has the optimal basic solution

$x^*_{10} = 0.6021$, $x^*_{20} = 0.1753$, $x^*_{31} = 0.0847$, $x^*_{41} = 0.0392$, $x^*_{51} = x^*_{61} = 0.0318$.

This yields an average cost optimal policy R* = (0, 0, 1, 1, 1, 1) with the minimal average cost $\sum_{i=2}^{N+1} x^*_{i1}$ = 0.206, in agreement with the results obtained by the policy-iteration algorithm.

Conclusion

The objective of the linear program is the minimization of the long-run average cost of treatment per unit time and the fraction of time in the long-run that a member staff could be in a poor condition of health and perhaps stays away from work.

This could be determined for each staff, so that for the staff whose value is a large contrast to that of the staff of the company could be considered as being in poor health condition quite often and therefore unproductive and may be retired. The cost obtained is not very realistic; it could be determined by other methods. The linear programming formulation has the advantage that sophisticated linear programming codes with the additional option of sensitivity analysis are widely available. The policy-iteration formulation usually involves the writing of its own code. However, the two methods are very efficient.

References

Abubakar, U. Y. (2011). Markov decision model for human health with policy iteration. *Journal of Science, Technology and Mathematics (JOSTMED), 7(3), 109-117.*

Abubakar, U. Y. (2005). A computer implementation of the revised simplex algorithm. ABACUS; *The Journal of the Mathematical Association of Nigeria. 32(2A), 24 – 34.*

Denardo, E. V., & Fox, B. L. (1968). Multichain markov renewal programs. *SIAM J. Appl Math. 16, 468-487.*

Diego, B., German, R. (2006). Linear programming solvers for markov decision processes. www.sys.virgina.edu|sieds06|papers|FmorningSession5.1. Date accessed: Jan.3rd 2011.

Goto, G. H., Lewis, M. E. & Puterman, M. L. (2004). Coffee, tea, or ...? A markov decision process model for airline meal provisioning. *Transportation Science, 38(1), 107-118.*

Hillier & Lieberman (1980). *Introduction to operations research.* NY: Holden Day.

Kulkani, V. G. (1999). *Modelling, analysis, design, and control of stochastic system.* Springer.

Puterman, M. (1994). *Markov decision processes: Discrete stochastic dynamic programming.* New York: John Wiley.

Tijm, H. C. (1988). *Stochastic modeling and analysis: A computational approach.* New York: John Wiley & Sons.