

## Markov Decision Model and the Application to the Cost of Treatment of Leprosy Disease

Usman Yusuf ABUBAKAR<sup>1</sup>, Sunday A. REJU<sup>2</sup> and Bamidele O. AWOJOYOGBE<sup>3</sup>

<sup>1</sup>*Department of Mathematics and Computer Science, Federal University of Technology, Minna, Nigeria*

<sup>2</sup>*National Open University, Lagos, Nigeria*

<sup>3</sup>*Department of Physics, Federal University of Technology, Minna, Nigeria*  
[abbkruy@yahoo.com](mailto:abbkruy@yahoo.com)

### Abstract

A Markov decision model is applied to determine the optimal cost of drugs in the treatment of leprosy disease. The result shows that the minimum cost is obtained if the Doctor is visited and the patient is placed on the MDT instead of traditional treatment and Dopsone monotherapy. The paper presents a theoretical framework to study the cost of treatment of leprosy where the relevant data could be obtained.

### Keywords

Markov Chain, Decision; Leprosy; Treatment; Cost; Dopsone - Monotherapy; Multi-Drug Therapy

### Introduction

Leprosy is an infectious disease of the skin and nerves caused by *mycobacterium Leprae*. It is wide spread in the tropical and sub tropical countries including Nigeria. Leprosy can be treated and cured with the Multi-Drug Therapy (MDT) or Dapsone-Monotherapy even without the associated deformity if it was detected early [1,2]. Prevalence rate and incidence of new cases of leprosy has be studied [3]. In addition, kinship coefficient has been used to study leprosy rates in villages of different distance apart [4]. In a related study, [5] has studied

the relapses in leprosy after treatment with Dapsone-monotherapy and Multi-Drug-Therapy (MDT) respectively. In this paper, we determine the best alternative in terms of drug-type in the treatment of leprosy disease.

### The Model Equations

Let  $X_n$  denote the state of the process at time  $n$  and  $a_n$  the action chosen at time  $n$ , then the above is equivalent to stating that:

$$P(X_{n+1}=j|X_0,a_0, X_{1,a_1},\dots,X_n=i, a_n=a) = P_{ij}(a) \quad (1)$$

This is the transition probability of a visit to state  $j$  given that state  $i$  was last visited when action  $a$  was chosen. Thus the transition probabilities are dependent on the present state and subsequent action [6].

A policy; by a policy we mean a rule for choosing actions. A policy is a sequence of decisions, one for each state of the process. It is denoted by  $d_i^{(n)}$  where  $n = 1, 2, \dots$  represent time and  $i = 1, 2, \dots$  represent state.

Following [7], we consider an aperiodic, irreducible Markov chain with  $m$  states ( $m < \infty$ ) and the transition probability matrix:

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ P_{m1} & P_{m2} & \dots & P_{mn} \end{bmatrix}$$

With every transition,  $i$  to  $j$  associate a reward  $R_{ij}$ . If we let  $V_i^{(n)}$  be the expected total earnings (reward) in the next  $n$  transitions, given that the system is in state  $i$  at present. A simple recurrence relation can be given for  $\{V_i^{(n)}\}_{n=1}^{\infty}$  as follows:

$$V_i^{(n)} = \sum_{j=1}^m P_{ij} [R_{ij} + V_j^{(n-1)}] \text{ where } i = 1,2,3,\dots,m; n = 1,2 \dots \quad (2)$$

$$\text{Let } \sum_{j=1}^m P_{ij} R_{ij} = Q_i$$

Equation (2) can now be written as:

$$V_i^{(n)} = Q_i + \sum_{j=1}^m P_{ij} V_j^{(n-1)} \quad (3)$$

setting  $n=1,2,\dots$  we get

$$\begin{aligned} V_i^{(1)} &= Q_i + \sum_{j=1}^m P_{ij} V_j^{(0)} \\ V_i^{(2)} &= Q_i + \sum_{j=1}^m P_{ij} [Q_j + \sum_{k=1}^m P_{jk} V_k^{(0)}] = Q_i + \sum_{j=1}^m P_{ij} Q_j + \sum_{k=1}^m \sum_{j=1}^m P_{ij} P_{jk} V_k^{(0)} \\ &= Q_i + \sum_{j=1}^m P_{ij} Q_j + \sum_{k=1}^m \sum_{j=1}^m P_{ij} P_{jk} V_k^{(0)} = Q_i + \sum_{j=1}^m P_{ij} Q_j + \sum_{k=1}^m P_{ik}^{(2)} V_k^{(0)} \end{aligned}$$

where  $P_{ij}^{(n)}$  is the  $(i,j)^{\text{th}}$  element of the matrix  $P^n$

$$\text{Let } V^{(n)} = \begin{bmatrix} V_1^{(n)} \\ V_2^{(n)} \\ V_3^{(n)} \\ \cdot \\ \cdot \\ V_m^{(n)} \end{bmatrix} \quad Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ \cdot \\ \cdot \\ Q_m \end{bmatrix}$$

Equation (3) can be put in matrix notation as

$$V^{(2)} = Q + PQ + P^2V^{(0)}$$

Extending this to a general  $n$ , we have

$$V^{(n)} = Q + PQ + P^2Q + \dots + P^{(n-1)}Q + P^nV^{(0)} = \left[ I + \sum_{k=1}^{n-1} P^k \right] Q + P^nV^{(0)}$$

### Theorem 1

Let  $p_{ij}^{(n)}$  be the  $n$ -step transition probability of an aperiodic, irreducible,  $m$ -state Markov chain. Let  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j$  ( $j=1, 2, 3, \dots, m$ ) be the limiting probabilities. Then there exist constants  $c$  and  $r$  with  $c > 0$  and  $0 < r < 1$ , such that  $p_{ij}^{(n)} = \pi_j + e_{ij}^{(n)}$  where  $|e_{ij}^{(n)}| \leq cr^n$

### Proof

From [7], we know that

$$a_n \leq P_{ij}^{(n)} \leq b_n \quad \text{and hence} \quad |e_{ij}^{(n)}| \leq b_n - a_n = d_n$$

From [7], we also have  $d_{kN+1} \leq (1 - 2\varepsilon_N)^k$ ; setting  $n = kN + 1$ , we get  $k = (n - 1)/N$  and

$$d_n \leq (1 - 2\varepsilon_N)^{(n/N) - (1/N)} \leq (1 - 2\varepsilon_N)^{n/(N-1)} = (1 - 2\varepsilon_N)^{-1} [(1 - 2\varepsilon_N)^{1/N}]^n$$

The theorem now follows by setting  $c = (1 - 2\varepsilon_N)^{-1}$  and  $r = (1 - 2\varepsilon_N)^{1/N}$ .

Consequently,  $P_{ij}^{(n)} = \pi_j + e_{ij}^{(n)}$

Where  $\{\pi_j\}_{j=1}^m$  is the limiting distribution of the Markov chain and  $|e_{ij}^{(n)}| \leq cr^n$  with  $c >$

0 and  $0 < r < 1$ . Therefore, as  $n \rightarrow \infty$ ,  $e_{ij}^{(n)} \rightarrow 0 \rightarrow 0$  geometrically.

$$\text{Let } \|e_{ij}^{(n)}\| = \eta(n) \text{ and } \Pi = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_m \\ \pi_1 & \pi_2 & \dots & \pi_m \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \pi_1 & \pi_2 & \dots & \pi_m \end{bmatrix}$$

In Matrix notation, we can write  $P^n$  as  $P^n = \Pi + \eta^{(n)}$

Substituting this in (3) we get:

$$V^{(n)} = \left[ I + \sum_{k=1}^{n-1} (\Pi + \eta^{(k)}) \right] Q + (\Pi + \eta^{(n)}) V^{(0)} = \Pi V^{(n)} + \left[ \sum_{k=1}^{n-1} \eta^{(k)} \right] Q + n \Pi Q + \eta^{(n)} V^{(0)} \quad (4)$$

In deriving (4), we have noted that  $\eta^{(0)} = 1 - \Pi$

Now consider the sum

$$\sum_{k=0}^{n-1} \eta^{(k)} = \sum_{k=0}^{\infty} \eta^{(k)} - \sum_{k=n}^{\infty} \eta^{(k)} \quad (5)$$

It should be noted that each term in  $\eta^{(k)}$  is less than or equal to  $cr^k$  ( $c > 0, 0 < r < 1$ ) in absolute value, and hence for large  $n$  the second term on the right hand side of (5) approaches an  $m \times m$  matrix with zero elements. For the same reason the last term in (4) approaches a null matrix for large  $n$ .

Thus asymptotically we have

$$V^{(n)} = \Pi V^{(0)} + \sum_{k=0}^{\infty} \eta^k Q + n \Pi Q$$

which gives

$$V_i^{(n)} = \sum_{j=1}^m \pi_j V_j^{(0)} + \sum_{j=1}^m \gamma_{ij} Q_j + n \sum_{j=1}^m \pi_j Q_j \quad (6)$$

where we have written  $\sum_{k=0}^{\infty} \eta^{(k)} = \|\gamma_{ij}\|$

writing

$$\sum_{j=1}^m \pi_j V_j^{(0)} + \sum_{j=1}^m \gamma_{ij} Q_j = \beta_i \quad \text{and} \quad \sum_{j=1}^m \pi_j Q_j = q$$

so that (6) can be put in the form

$$V_i(n) = \beta_i + nq$$

which shows that for large  $n$ ,  $V_i^{(n)}$  is a linear function of  $q$  for every  $i$ . Further, for different values of  $i$ ,  $V_i^{(n)}$  are represented by parallel straight lines with slope  $q$  and intercept  $\beta_i$  ( $i = 1, 2, \dots, m$ )

So far, we have considered the transition probability matrix  $P$  and the reward matrix  $R$  as given.

Instead, suppose that the decision maker has other alternatives and so is able to alter elements of  $P$  and  $R$ . To incorporate this feature, we define  $D$  as the decision set, and under decision  $K \in D$ , let  ${}^K P_{ij}$  and  ${}^K R_{ij}$  be the probability of the transition from  $i$  to  $j$  and the corresponding reward respectively for  ${}^K V_i^{(n)}$ . The expected total earnings in  $n$  transitions under decision  $K$ ; we have the recurrence relations ( $K=0$  represents the optimal decision)

$${}^0 V_i^{(n)} = \max_{k \in D} \sum_{j=1}^m {}^k P_{ij} [{}^k R_{ij} + {}^0 V_j^{(n-1)}] \quad n = 1, 2, \dots, i = 1, 2, \dots, m \quad (7)$$

giving

$$V_i^{(n)} = \max_{k \in D} \left[ {}^k Q_i + \sum_{j=1}^m {}^k P_{ij} {}^0 V_j^{(n-1)} \right]; \quad i = 1, 2, \dots, m, n = 1, 2$$

where we have written  $\sum_{j=1}^m P_{ij} {}^k R_{ij} = {}^k Q_i$

Recursive relation (7) gives an iteration procedure to determine the optimum decision  $d_i^{(n)} \in D$ . For  $i = 1, 2, \dots, m$  and  $n = 1, 2, \dots$

This is a Standard technique in Dynamic programming and it has been shown Bellman [8,9], that this iteration process will converge on best alternative for each state as  $n \rightarrow \infty$ .

The method is based on recursively determining the optimum policy for every  $n$ , that would give the maximum value.

However, one major drawback of the method is that, there is no way to say when the policy converges into a stable policy; therefore, the value iteration procedure is useful only when  $n$ , is fairly small

### Application

We consider a new case of leprosy, that is, a person showing clinical signs of leprosy, with or without bacteriological confirmation of the diagnosis (state 1), or a new case with WHO disability grade 1 (state 2), or a new case with WHO disability grade 2 (state 3) [1,2]. Thus, we have a three states model. We assumed that all the states communicate (Ergodic). Let the transition between the states be represented by the transition diagram shown in figure 1, governed by the following transition probabilities of Markov chain in (1) and Markov reward matrices respectively.

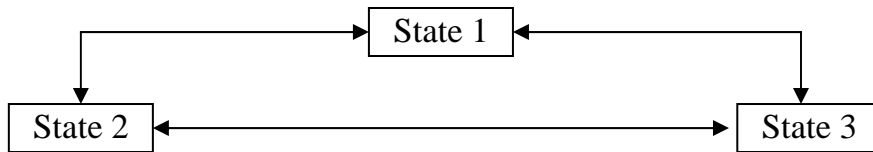
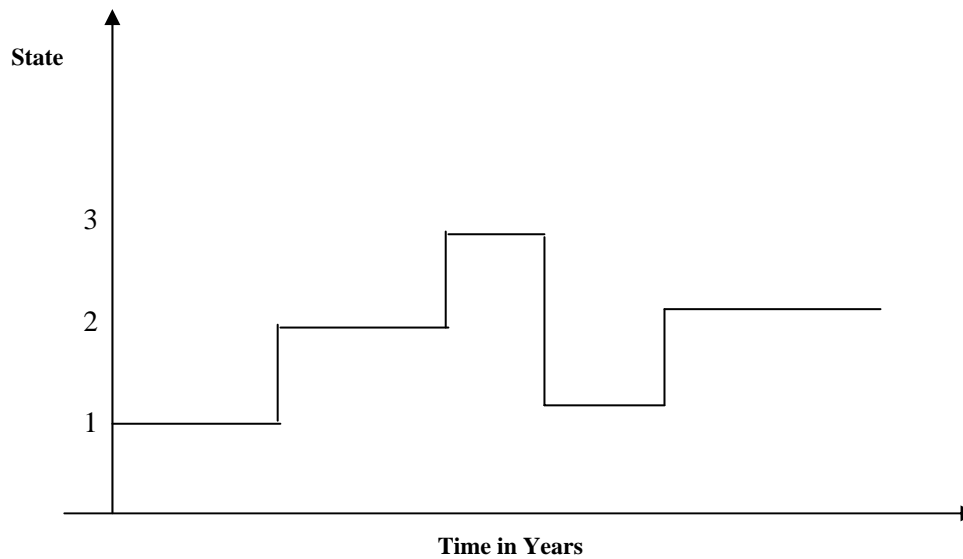


Figure 1. The Transition Diagram for Leprosy Patients

$$P = (P_{ij}) = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}; i, j, = 1, 2, 3 \text{ where } 0 \leq p_{ij} \leq 1 \text{ and } \sum_{j=1}^4 P_{ij} = 1, i = 1, 2 \text{ and } 3 \text{ and}$$

$$R = (R_{ij}) = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}; i, j = 1, 2, 3.$$

Figure 2 presented the possible realization of the state transitions.



*Figure 2. A possible realization of the state transitions*

At any given point in time and state of the disease; the patient has the opportunity or a privilege to make a choice of drugs. The choice of drugs may however be dependent on the availability of the drugs, resources available to the patient and the state or severity of the disease. The two major types of drugs used for the treatment of leprosy are Dapsone-Monotherapy and Multi-Drug therapy (MDT) [1,2] respectively.

For the sake of simplicity, we may further divide the drugs into Low priced drugs and high priced drugs to correspond to Dapsone and MDT respectively. We shall have the following alternatives for the states [10]:

- **State1:**
  - Alternative 1: Self-medication (Traditional treatment)
  - Alternative 2: Go to see a doctor (cost of seeing a Doctor).
- **State 2:**
  - Alternative 1: Low priced drugs
  - Alternative 2: high price drugs
- **State 3:**
  - Alternative 1: Continue without change of drugs
  - Alternative 2: Change drugs (from low to high priced or vice versa).

Costs are associated with each of these alternatives.

Our objective is to obtain the policy or alternative that minimizes the costs of the treatment at any given state and time.

Let the transition probabilities ( $P_{ij}$ ) and the corresponding reward ( $R_{ij}$ ) be given as follows:

$$P = (P_{ij}) = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}; i, j = 1, 2, 3$$

and

$$R = (R_{ij}) = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}; i, j = 1, 2, 3$$

Let  $D$  be the decision set and we have two Alternative decisions available to the patient. That is, Alternative 1; and Alternative 2; Thus in every state we have  $k = 1, 2 \in D$ .

We shall now determine the best policies for every  $n$  using equation (7).

Since we are concerned about minimizing the costs of drugs to be administered; the alternative that yields the least value constitutes the best policy for the state and time.

## Results

Suppose we have the following stationary transition probabilities and the corresponding cost matrices.

$$P = \begin{pmatrix} 0.795 & 0.077 & 0.128 \\ 0.235 & 0.441 & 0.324 \\ 0.095 & 0.190 & 0.714 \end{pmatrix}$$

and

$$R = \begin{pmatrix} 0 & 1 & 3 \\ 2 & 3 & 4 \\ 1 & 2 & 4 \end{pmatrix} \text{ in } \text{N}1000 \text{ respectively}$$

The subsequent transition probabilities and cost matrices are given below when the decisions are implemented.

$$P_{i,j} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{pmatrix} \quad \text{and} \quad R_{ij} = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 2 & 3 \\ 1 & 2 & 4 \end{pmatrix}$$



$i, j = 1, 2, 3 \text{ for } k = 1$ 

$$P_{i,j} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.6 & 0.3 & 0.1 \\ 0.5 & 0.4 & 0.1 \end{pmatrix}$$

 $i, j = 1, 2, 3 \text{ for } k = 1$ 

$$\text{and } R_{ij} = \begin{pmatrix} 1 & 2 & 2 \\ 3 & 2 & 4 \\ 3 & 2 & 1 \end{pmatrix}$$

 $i, j = 1, 2, 3 \text{ for } k = 2$ 
 $i, j = 1, 2, 3 \text{ for } k = 2$ 

We shall use these values to determine the best policies for each  $n$ .

The summary of the results is presented in Table1.

Table 1. A Summary result of the Optimal Policies and Costs

$n$	$d_1^{(n)}$	$d_2^{(n)}$	$d_3^{(n)}$	${}^oV_1^{(n)}$	${}^oV_2^{(n)}$	${}^oV_3^{(n)}$
1	2	1	2	1,200	2,300	2,400
2	2	2	2	2,850	4,450	4,160
3	2	2	2	4,340	6,260	6,020
4	2	2	2	5,900	7,880	7,680
5	2	2	2	7,480	9,470	9,270

## Discussion

The results indicate the best policies for each  $n$ .  $d_i^{(n)}$  where  $n = 1, 2, 3, 4, 5$  and  $i = 1, 2, 3$ . Thus, we have obtained the best policies for the three states for five years. In addition to the best policies, the corresponding expected minimum costs are also provided. For instance,  $d_1^{(1)} = 2$  with  ${}^oV_1^{(1)} = 1,200$  means that the best policy for state 1 for the first year is to see the doctor and the corresponding expected minimum cost is one thousand and two hundred Naira (₦ 1200.00).

The results revealed that except for the  $d_2^{(1)} = 1$  with the corresponding  ${}^oV_2^{(1)} = 2,300$ , the best policy for the other states at each time is the alternative 2. Which means that the best policy for each other state at each time is the second alternative. This is a kind of convergence to a stable policy. The convergence is so smooth that further iteration beyond five years is not necessary. This type of convergence is not generally true of this iterative algorithm, however, it makes this illustration very interesting.

## References

---

1. WHO, WHO Expert Committee on Leprosy. Seventh report, Technical report Series 874, 1997.
2. ILEP, ILEP Medical Bulletin No 14 Operational Guidelines for the Introduction of new MDT Regimens for the Treatment of leprosy, 1998.
3. Lechat M. F., The torments and blessings of the Leprosy epidemometric model, *Leprosy Review*, 1981, 52, p. 187-196.
4. Bechelli L. M., Barraï I., Gallego P. et all., Correlation between leprosy rates in villages different distances apart, *Bulletin World Health Organization*, 1973, 48, p. 257-378.
5. Abubakar U. Y., A Study of Relapses in Leprosy: A Semi-Markov Approach, *Abacus, The Journal of the Mathematical Association of Nigeria*, 2004, 31(2B), p. 370-377.
6. Ross S. M., *Introduction to Probability Models*. Academic Press, Inc. New York, 1989.
7. Bhat U. N., *Element of applied Stochastic Processes*, John Wiley New York, 1984.
8. Bellman R., *Dynamic Programming*. Princeton University Press, Princeton, 1957.
9. Bellman R., A Markov Decision Process, *J. Math. Mech.*, 1957, 6, p. 679-684.
10. Abubakar U.Y., *The Application of Markov and Semi-Markov Models to the Control of Catarrh and Leprosy diseases*, PhD Thesis (Unpublished), Federal University of Technology Minna, 2004.