Markov Decision Processes and the Modelling of Patient Flows

Anthony Clissold
Supervised by Professor Jerzy Filar
Flinders University

1 Introduction

Australia's growing and ageing population and a rise in chronic diseases are causing an increased demand for emergency and inpatient services in the country's hospitals (FitzGerald 2012). Currently, many hospitals are operating at or close to full capacity a lot of the time. Bed availability is randomly distributed because of the random nature of patient arrivals and the duration of treatment. Due to this, episodes of congestion can occur, which lead to wide-ranging consequences such as fatigued staff and decreased quality of care, amongst others. This demand is projected to increase into the future as the proportion of older Australians increases.

Despite the fact that bed availability is stochastic in nature, most hospitals experience regular patterns in their occupancy levels, on both day-to-day and week-to-week time scales. These patterns can be easily modelled mathematically as *Markov processes* (Qin, Filar 2014).

Definition 1.1: A finite Markov chain

A sequence of random variables $\{X_n\}_{n=0}^{\infty}$ is a finite Markov chain if it satisfies the following criteria:

- There are finitely many states $I = \{1,2, \dots, m\}$ and $X_n = x_n \in I$ at every n.
- The Markovian property is satisfied for all $j_0, j_1, \dots, j_{n-1}, j \in I$:
  $P(X_n = j | X_0 = j_0, \dots, X_{n-1} = j_{n-1}) = P(X_n = j | X_{n-1} = j_{n-1})$

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia
Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

The Markov chain is homogeneous if, in addition, the following property holds:

- The time homogeneity property: $P(X_n = j|X_{n-1} = i) = p_{ij}, \forall i, j \in I, n = 1,2, \ldots$

The Markovian property states that the probability of transitioning to a state is only dependent on the current state, but independent of all states visited previously. The time homogeneity property implies that the probability of a transition between any two states is independent of the number of time periods that have passed before the transition occurs. Hence, a homogeneous Markov chain is fully determined by its probability transition matrix $P = \left(p_{ij}\right)_{i,j=1}^{m,m}$.

Markov chains have been widely used to model a variety of stochastic phenomena where the current situation depends only on the current state (and possibly time); see for instance Heyman and Sobel (1984). However, in situations where the probability transition matrices may be influenced by a decision maker's (controller's) actions, the problem to be analysed becomes that of a choice of a "best" Markov chain to meet certain performance criteria. The latter usually involve maximising expected rewards or minimising expected costs. This class of models is known as Markov Decision Processes (MDP's for short) that have been studied extensively since Howard's original work in 1960. They can also be viewed as discrete, stochastic dynamic programming problems to which Bellman's (1957) backward recursion algorithm applies.

In this application, we shall use a given hospital's occupancy data to construct probability transitions consistent with these data. Also, we introduce a reward function that attempts to balance the benefit of maintaining sufficient slack in the occupancy of wards versus the cost of diverting patients elsewhere. This information can be given to staff in the hospital to change policies in the hospital's operation, thus achieving lower occupancy, higher quality of care, shorter length of stay and higher staff morale, plus many more flow-on benefits throughout the health-care system.

It should be mentioned that this study is preliminary, and its results will form a proof of concept of the Markov Decision Process approach to this problem.

2 Data and Modelling of Probability Transition Matrices

2.1 Data

The data used in this project was midnight census data at Flinders Medical Centre, Bedford Park, SA, from January 2009 to June 2013. The only patients included were General Medical and Surgical; these patients make up a large proportion of Flinders' inpatients, approximately 65% of all inpatients (SA Health, 2014). The week

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

between December 25th and 31st inclusive was removed from the analysis as current policies heavily minimise bed occupancy during this time. This allows staff to take leave over the Christmas-New Year period.

2.2 Modelling of daily probability transition matrices

The midnight census data can be modelled as a Markov chain by converting each census value into a pre-defined aggregate state and then modelling a probability transition matrix from state to state. The census data was approximately normally distributed as seen from Figure 2.2.1.

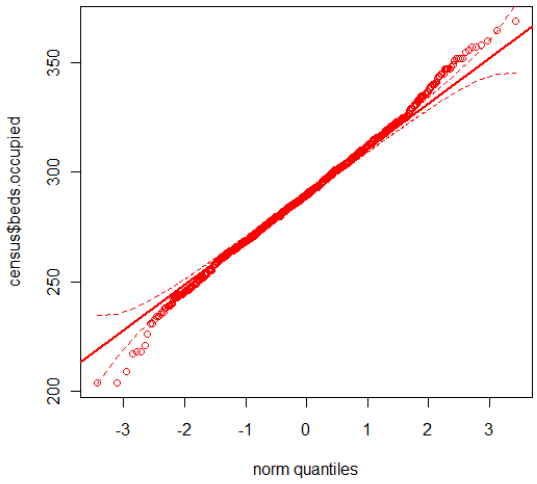

Figure 2.2.1: Normal Q-Q plot of Flinders Medical Centre midnight census data from January 2009 – June 2013

A simple, four state, approach has been used in this project as it gives a good balance between practice and theory. Too many states make the probability transition matrices too inaccurate (and difficult for managers to use), whereas too few make the model too coarse. Based on consultation with Flinders Medical Centre (FMC) staff, the four occupancy states were defined on the basis of deviations from the mean occupancy. The boundaries between the four states are given below.

Definition 2.2.1: Boundaries between states

The states {1,2,3,4} will correspond intuitively to Low, Medium, High and Very High occupancies as indicated below:

- Low ($S = 1$): $s \leq \mu - \frac{\sigma}{2}$
- Medium ($S = 2$): $\mu - \frac{\sigma}{2} < s \leq \mu + \frac{3}{4}\sigma$

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

- High ($S = 3$): $\mu + \frac{3}{4}\sigma < s \leq \mu + \frac{3}{2}\sigma$
- Very High ($S = 4$): $s > \mu + \frac{3}{2}\sigma$,

where the aggregated states denoted by S actually correspond to collections of micro-states denoted by s that are the daily occupancy levels.

As a consequence of the approximate normality of the occupancy data, the proportions of time spent in each state can be estimated. They are as follows: Low, 31%; Medium, 46%; High, 16% and Very High, 7%.

Each day can now be classified as a day of one of Low, Medium, High or Very High occupancy, and then the transition from day to day can then be classified as one of the 16 state-to-state pairs (Low-Low, Medium-High etc.)

Hence, there are seven daily probability transition matrices,$P_0, P_1, \ldots, P_5, P_6$ one for each day of the week. P_0 represents transitions from Monday to Tuesday, P_1 represents transitions from Tuesday to Wednesday and so on. Note that because these matrices, when extracted from data, are distinct the daily Markov chain is inhomogeneous. However, the Markovian property is generally well satisfied (see Section 2.4 below).

2.3 Daily probability transition matrices

One of the daily probability transition matrices derived from the data can be found in Table 2.3.1. The complete set can be found in Appendix A. The given matrix is for transitions from Monday to Tuesday (P_0).

0.8293	0.1707	0	0
0.1525	0.7966	0.0508	0
0	0.4694	0.4286	0.102
0	0.0385	0.6538	0.3077

Table 2.3.1: Monday to Tuesday probability transition matrix of Flinders Medical Centre midnight census data from January 2009 – June 2013

It can be seen from the above that the first two diagonal entries of P_0 are dominant. Thus the probability of remaining in a Low or Medium state on Tuesday as the one the system was in such a state on Monday is clearly the highest. However, for High and Very High states, there remain significant probabilities of either moving to or remaining in the Very High state. Arguably, the latter is an undesirable situation. A range of interesting patterns of that type can also be observed by examining the probability transition matrices for the remaining days. See Appendix A.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

2.4 Weekly probability transition matrices

It was found by Qin and Filar (2014) that weekly probability transition matrices $Q_0, Q_1, \ldots, Q_5, Q_6$ can be calculated as products of the daily probability transition matrices such that

$$Q_0 = P_0P_1 \ldots P_5P_6; \quad Q_1 = P_1P_2 \ldots P_6P_0; \quad \ldots \quad Q_6 = P_6P_0 \ldots P_4P_5. \quad (1)$$

Therefore, Q_0 represents weekly probability transitions from a Monday to the following Monday, Q_1 represents similar transitions from a Tuesday to the following Tuesday, and so on. It should be noted that these matrices observe the time homogeneity property, thus defining a homogeneous Markov chain.

Table 2.4.1 contains a Monday-to-Monday weekly probability transition matrix that was calculated from the daily probability transition matrices given in Appendix A. The full set of weekly probability transition matrices can be found in Appendix B.

0.2101	0.5405	0.1794	0.07
0.1837	0.5182	0.2012	0.0969
0.1539	0.4804	0.2281	0.1377
0.1282	0.4427	0.2521	0.177

Table 2.4.1: Monday-to-Monday probability transition matrix of Flinders Medical Centre midnight census data from January 2009 – June 2013

An important observation to be made when comparing daily and weekly probability transition matrices is that the latter contain no zero entries and hence are automatically irreducible. From discussions with hospital staff we inferred that in terms of management policies, it is the long-term behaviour induced by the weekly probability transitions that could influence these policies rather than any daily transitions.

2.5 Steady State Distribution

In a finite Markov chain, two states *communicate* if it is possible to reach one from the other in a finite number of transitions, and vice versa. A Markov chain is *irreducible* if it contains a single, exhaustive, class of communicating states. This is the case in each of the weekly transition matrices: each state communicates with all of the others, therefore they form a single exhaustive communicating class. These two conditions are sufficient for the unique steady state distribution to exist.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

Definition 2.5.1: Steady State Distribution

The steady state distribution, π, of a finite, irreducible Markov chain is given by the unique solution to the following equations:

$$\pi P = \pi, \qquad \sum_i \pi_i = 1.$$

The resulting vector, π, gives the probability of being in any particular state as $t \to \infty$ (the steady state probability).

Let $\pi_0, \ldots, \pi_6$ denote the steady state distributions of $Q_0, \ldots, Q_6$. We shall use the notation $\pi_i = (\pi_{i1}, \pi_{i2}, \pi_{i3}, \pi_{i4})$ where π_{34} denotes the steady state probability of the Very high state on Thursday.

Figure 2.5.1 shows the trends of these weekly steady state probabilities. For instance, the above mentioned π_{34} is depicted as the fourth point on the black curve in that figure.

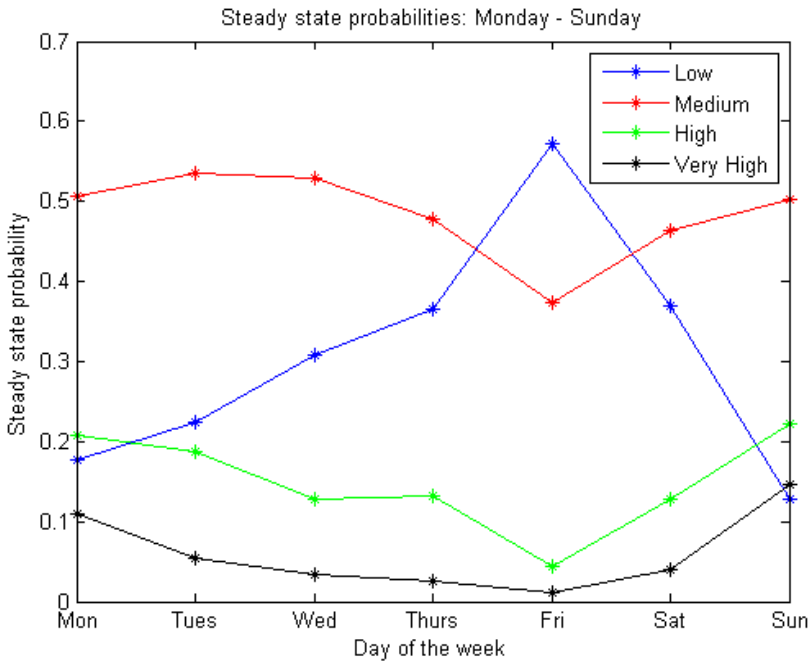

Figure 2.5.1: Comparison of steady state probabilities for each day of the week

Figure 2.5.1 shows that as the week progresses, the likelihood of being in a higher occupancy state decreases, reaching a minimum on Fridays. This coincides with the discharge of patients for the approaching weekend. On weekends, the probability of being in a higher state begins to increase again.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

2.6 Validation of the Model

It is important to validate the transition matrices to account for any major deviations from the observed data. There is a rule of thumb to use the first two-thirds of the data for parameter estimation, and the remaining one-third for validation. We chose to do the validation on the basis of fit of the steady state distributions. The steady state distributions, $\pi_0, \ldots, \pi_6$, were found, as above, for each weekly probability transition matrix. For the remaining one-third of the data, the proportions of weekly occupancy lying in states Low, Medium, High and Very High were calculated for each day of the week, resulting in seven vectors $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}),\ \ i = 0, \ldots, 6$. The discrepancy between the fitted model and the data was measured by the index:

$$d_i = \frac{1}{4}\sum_k \left[\frac{\pi_{ik} - x_{ik}}{\pi_{ik}}\right]^2 \quad i = 0,1, \ldots, 6 \qquad (2)$$

where π_{ik} is the steady state probability of being in state *k* for the fitted probability transition matrix and x_{ik} is the corresponding proportion for the remaining one-third of the data. The index d_i is the mean squared deviation error. When it is small, it can be argued that the fitted parameters are adequate. The calculated d_i values are given below:

d_0	0.1552
d_1	0.0588
d_2	0.0347
d_3	2.3991
d_4	0.1142
d_5	0.1656
d_6	0.0581

Figure 2.6.1: Calculated *d* values for each weekly stationary distribution

For six out of the seven daily transitions, the calculated *d* values were, indeed, small (ranging from 0.04 to 0.17). For the remaining one, Thursday to Friday, the calculated *d* value was 2.3991. While this is high, the steady state vector $\pi_3 = (0.3505,\ 0.4657, 0.1741, 0.0097)$ was being compared to the observed frequencies of occupancies being in the Low, Medium, High and Very High states in the final third used for validation. The latter was $x_3 = (0.3926, 0.4701,\ 0.098, 0.0392)$. Thus, it is only the deviation in the low probability Very High state that contributes most to this high index value. In view of this, we decided that for the bulk of the estimated parameters the modelled probability transitions were adequate.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

3 Derivation of the Markov Decision Process

Next, we cast the congestion relief problem in the framework of a finite horizon Markov Decision Process. Towards that goal, we need to define probability transitions that depend on decision maker's actions and the immediate "rewards" to the decision maker resulting from a choice of a given action in a given state. We adopt notation similar to that used in Filar and Vrieze (1997).

Definition 3.1: Markovian Transition Probabilities and Rewards of the MDP

In a finite Markov Decision Process, the system is in state $s \in S = \{1,2,\dots,N\}$ at time *t*, and the decision maker is obliged to choose an action $a \in A(s) = \{1,2,\dots,m_s\}$. As a result of this choice, the following stochastic transition occurs:

- $p(s,s',a) = P(S_{t+1} = s' | S_t = s, A_t = a)$,

and an immediate reward (cost if negative) is accrued

- $r(s,a)$.

Note that, in some states, there could be only a single action that corresponds to a decision "no action needed". Indeed, this is the case in the Low and Medium occupancy states in our model. To fully describe the MDP model we need to specify the set *S* of states, the sets *A(s)* of actions available in each state, the time horizon *T* and the above probability transitions and rewards. These will be called the parameters of the MDP.

3.1 MDP parameters

In our four state occupancy model, actions will require physical interpretation that correspond to various levels of intervention designed to lower the occupancy. Some of the parameters used in the original probability transition matrices will stay the same in the corresponding MDP to be developed, such as the states and their definitions. Also, it should be noted that the transition probabilities will depend on the actions taken only to the extent that the actions taken would have moved the system from one state to a lower state if nothing else occurred. This is because the stochasticity in the model is due to external factors beyond the decision maker's control. The latter can be taken advantage of because the probability transitions of the MDP model can be easily derived from the historical data of the hospital's occupancy.

3.1.1 Actions

Many possible actions could be taken to mitigate congestion episodes in a hospital. They include discharging patients early, cancelling elective treatments or transferring some patients to other hospitals. All of these actions involve removing some number of

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

patients from the system, so each possible action will be defined as removing a multiple of four patients at a time from the system up to some practical upper bound. In this proof of concept exercise we have decided that the set of all possible actions will be:

- 0: No action
- 1: Remove four patients
- 2: Remove eight patients
- 3: Remove 12 patients
- 4: Remove 16 patients
- 5: Remove 20 patients.

Twenty was chosen as a cap for removal of patients, as it seemed that removing any more patients than that was impractical when considered on top of the current policies.

We have also decided that $A(1) = A(2) = \{0\},\ A(3) = A(4) = \{1,2, \dots ,5\}$. Thus, the patient lowering actions can be invoked only in the High and Very High states.

3.1.2 Rewards

Next, the reward functions $r(s, a)$ need to be modelled. The challenge here is to capture both the benefit of reducing occupancy when it is too high and the cost resulting from inconveniencing patients and the associated harm to the reputation of the hospital. Even though these are qualitative benefits and costs, we require a quantitative expression that balances them.

Since our states are labelled 1,2,3,4 in the order of degree of occupancy - and the actions are also labelled in the order of increasing numbers of patients being removed - it is natural to assume that the relief provided by a removal of a certain number of patients is more significant in the Very High state than in the High state. The cost, on the other hand, is assumed to increase exponentially with the number of patients removed from the system. The following reward functions involve three nonnegative parameters $\alpha, \beta, \gamma, \delta$ whose role is to capture an appropriate trade-off between the above cost and benefit. The $20 - 4a$ term in the first equation enables the High state's reward function to peak at a more reasonable position when compared to the corresponding function for the Very High state. The following forms of the function have been adopted for each state:

$$r(3, a) = \delta\left[\left(\frac{20-4a}{369} \times 100\right) 3^{\alpha} - \gamma e^{\beta(20-4a)}\right];\ \ a \in A(3). \qquad (3)$$

$$r(4, a) = \ \delta\left[\left(\frac{4a}{369} \times 100\right) 4^{\alpha} - \gamma e^{4\beta a}\right];\ \ a \in A(4). \qquad (4)$$

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

The first term of (3) and (4) represents the benefit of the increase in bed space by executing action a, the second term quantifies the loss of reputation for the hospital for removing $4a$ patients. The parameter β must be small enough so that the exponential term does not dominate the entire expression. A method of numerical calibration was used to determine the parameters needed for the function. For the states $s \in \{1,2\}$, namely, the Low and Medium states, the only available action is action 0, and hence the immediate rewards were selected to be constants representing a small benefit from being in a desirable state. The resulting $r(s,a)$ functions used in our MDP model are:

$$r(1,0) = 0.8 \tag{5}$$

$$r(2,0) = 0.7 \tag{6}$$

$$r(3,a) = \frac{1}{7}\left[\left(\frac{20-4a}{369} \times 100\right) 3^{0.8} - \frac{2}{3} e^{0.15(20-4a)}\right];\ \ a \in A(3). \tag{7}$$

$$r(4,a) = \frac{1}{4}\left[\left(\frac{4a}{369} \times 100\right) 4^{0.3} - 0.3e^{0.64a}\right];\ \ a \in A(4). \tag{8}$$

The following figure shows a plot of the values of the reward functions against the number of patients removed for the High and Very High states. The two remaining states have a single possible action; hence they are excluded from the plot.

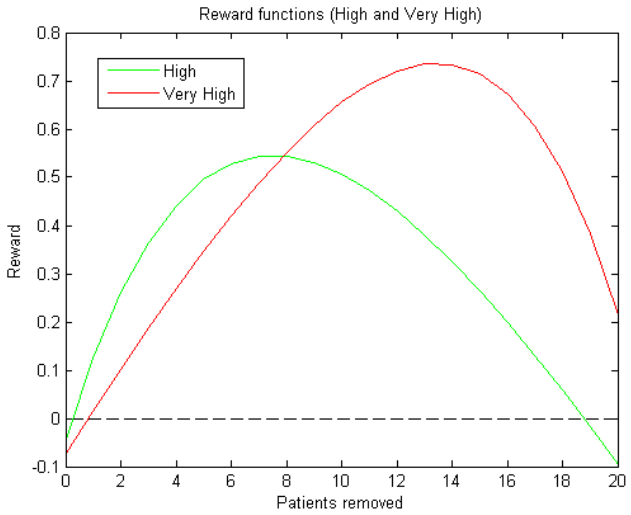

Figure 3.1.2.1: Plot of the High and Very High reward functions, $r(3,a)$ and $r(4,a)$, for zero to 20 patients removed

The reward function, $r(4,a)$, for the Very High state begins slightly negative (indicating a slight penalty for no action), rises to a peak when approximately 14 patients are removed after which it decreases. The reward function for being in the High state, $r(3,a)$, also begins slightly negative, but peaks when eight patients are removed. It then decreases and becomes a cost (negative reward) when more than 19 patients are removed.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

The maximum reward that can be accrued when in the Very High state is greater than that for High because freeing up extra capacity carries a greater premium when there are fewer beds available. This also allows for a larger decrease in occupancy before the cost of inconveniencing patients overcomes the benefit of increased capacity. The reward accrued when fewer patients are removed is greater when in the High state in comparison to the Very High state as this has less effect in alleviating the extreme congestion encountered in the Very High state, this is why the green curve lies above the red curve in the range [0,8). Similarly, it is natural that the green curve lies below the red curve for higher values of a.

3.2 Finding Probability Transition Matrices under Actions

The actions defined in a Markov Decision Process alter the probabilities of a transition from one state to another. There needs to be a simple way to estimate these probabilities. One method of doing this is by splitting the aggregated states into micro-states for each number of beds occupied. Let $n_{s,s'}$ be the number of observed transitions from s beds occupied on one day to s' beds occupied on the following day, and N_s be the total number of days that s beds were occupied, thus:

$$\hat{P}(X_{t+1} = s' | X_t = s) = \frac{n_{s,s'}}{N_s} \tag{9}$$

This is valid without considering the effects of any actions on these probabilities. When action a is taken, the transition effectively becomes a transition from $s - 4a$ beds occupied to s' beds occupied, which is estimated by:

$$\hat{P}(X_{t+1} = s' | X_t = s, A_t = a) = \frac{n_{s-4a,s'}}{N_{s-4a}} \tag{10}$$

Using this equation it is now possible to estimate the new transition probabilities of moving between the aggregated states under a given action. By the rules of calculus of probabilities, these were found to be:

$$P(X_{t+1} \in S' | X_t \in S, A_t = a) = \tag{11}$$

$$= \sum_{s' \in S'} \sum_{s \in S} P(X_{t+1} = s' | X_t = s, A_t = a) P(X_t = s | X_t \in S)$$

The first factor in each of the terms of the double summation in (11) is equal to the probability of moving from each micro-state in aggregated state at time t to each micro-state in the aggregated state at time $t + 1$, conditioned on the micro-state at time t being s and the action chosen being a. The second factor is the probability of the micro-state at time t being s conditioned on the macro-state being S. This second factor will be estimated with the help of steady state probabilities in a natural way. This leads to the following formula:

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

$$P(X_{t+1} \in S' | X_t \in S, A_t = a) = \sum_{s' \in S'} \sum_{s \in S} P(X_{t+1} = s' | X_t = s, A_t = a) \frac{\pi_s}{\sum_{i \in S} \pi_i} \quad (12)$$

By substituting (10), the above formula can be further simplified to:

$$\hat{P}(X_{t+1} \in S' | X_t \in S, A_t = a) = \sum_{s' \in S'} \sum_{s \in S} \frac{n_{s-4a,s'}}{N_{s-4a}} \frac{\hat{\pi}_s}{\sum_{i \in S} \hat{\pi}_i} \quad (13)$$

The use of actions in the above formula presents a technical challenge, as the raw data matrix of observed transitions ($n_{s,s'}$) is quite sparse. This means that some rows that will be needed in the computation of the probabilities are empty.

To alleviate this problem, the empty rows of the matrix were replaced with estimates. We chose to replace each empty row with the closest non-empty row. All entries were rounded up to the nearest integer, as a fraction of a transition count is inappropriate. Once all empty rows have been removed, the calculations for each of the new transition probabilities are well defined.

3.2.1 Steady state distribution of the sparse matrix

Another difficulty with using such sparse probability transition matrices lies in calculating their steady state distributions. Because some transitions are not observed on a particular day and the sparseness of the matrix results in multiple communicating classes, the unique steady state distributions may not exist.

However, the raw data transition matrix may be re-constructed in such a way that there is only one communicating class. This can be done by constructing a matrix with every transition in the data set and then adding a virtual transition from the last micro-state observed back to the first one. This amounts to a small perturbation of the raw data transition matrix. Now every observed micro-state can be reached from every other observed micro-state, which enables the unique steady state distribution to be calculated for each micro-state on each day of the week. In the constructed matrix, each set of transitions for each day of the week occupies its own unique block of rows and columns that make it easy to extract the steady state distribution from the matrix. The layout of this matrix is shown in Figure 3.2.1.1.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

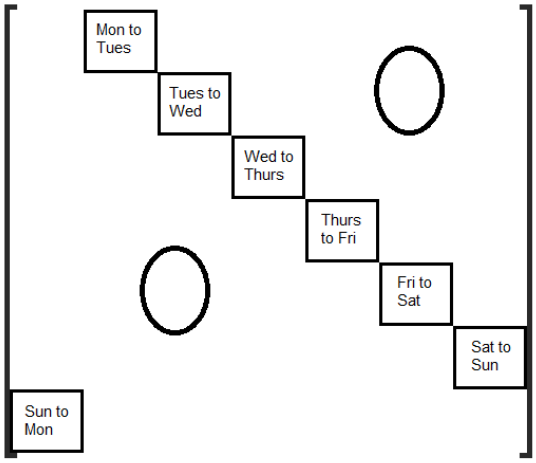

Figure 3.2.1.1: Diagram of the sparse matrix constructed to calculate the steady state distribution for each day of the week

The steady state vector for the constructed matrix is made up of seven sections, each corresponding to a day of the week. The corresponding sections for each day can be removed from the larger vector and normalised to form the steady state distribution vector for a particular day of the week.

Now that the steady state distribution has been calculated, it is possible to estimate the daily probability transition matrices for each day of the week under any given action *a*. The Markov Decision Process is fully defined once the actions, reward functions and probability transition matrices are known.

4 Finding an Optimum Policy for the MDP

With a fully defined MDP, the next task is to find the optimum *policy* for the given process. A policy is a sequence of actions that defines what happens to the MDP on a given time horizon. Strictly speaking, the preceding only defines a deterministic Markov policy, but these are sufficient to find optimal policies even in larger policy spaces that are not of interest here (e.g. see Filar and Vrieze (1997)).

Using a finite time horizon, the standard method for calculating optimal policies in an MDP is the Dynamic Programming algorithm. This can be very time consuming due to the well-known phenomenon known as “the curse of dimensionality”. The latter is not too much of a problem when the time horizon is short and the numbers of states and actions are small. Fortunately, the MDP in our application is sufficiently small that the computational burden is reasonable. A time horizon of seven days will be used in this analysis to replicate the week-long planning patterns at Flinders Medical Centre.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

4.1 The Backward Recursion of Dynamic Programming Algorithm

As mentioned above, the main algorithm to find an optimal policy is the Backward Recursion of Dynamic Programming algorithm, which was developed in 1957 by Bellman. The notation used below is the notation from Filar and Vrieze (1997). The Backward Recursion algorithm is a recursive algorithm that works as follows:

Step 1 (Initiation): Set $V_{-1}(s) = 0 \quad \forall s \in S$ and define

$$f_T^*(s) \coloneqq a_s^T = argmax_{A(s)}\{r(s,a) + \sum_{s'=1}^{N} P_T(s'|s,a)V_{-1}(s)\}$$

$$V_0(s) \coloneqq r(s, a_s^T) = max_{A(s)}\{r(s,a)\}$$

Set the expected reward for being in each state to zero and then define the optimal action for each state as the one which maximises the immediate reward. In this step, the summation term collapses to zero. Finally, set the expected reward for each state to the maximum immediate reward.

Step 2 (Recursion): For each $n = 1,2,\dots,T$ and each $s \in S$

$$f_{T-n}^*(s) \coloneqq a_s^{T-n} = argmax_{A(s)}\{r(s,a) + \sum_{s'=1}^{N} P_{T-n}(s'|s,a)V_{n-1}(s)\}$$

$$V_n(s) \coloneqq r(s, a_s^{T-n}) + \sum_{s'=1}^{N} P_{T-n}(s'|s,a)V_{n-1}(s)$$

Working backwards in time, for each state in the state space, this step of the algorithm defines an optimal action to be taken at that time.

Step 3: Construct a policy

$$f^* = [f_0^*, f_1^*, \dots, f_T^*]$$

By putting together the actions which maximise the expected reward for each state and time point combination, the optimal policy is found.

The validity of the algorithm depends on Bellman's principle of optimality, which is summarised well for this application by Filar and Vrieze (1997):

> 'If an optimal reward can be found for the process with (n-1) steps left, then the optimal reward can be found for n steps left by maximising the sum of the immediate reward and the maximal reward for the process with (n-1) steps left.'

The proof that f^* is in fact the optimal policy is given on page 20 of Filar and Vrieze (1997).

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

4.2 The Optimal Solution to the MDP

Running the Backward Recursion of Dynamic Programming algorithm on the Markov Decision Process defined earlier yields an optimal solution for each state and time point as follows, where the columns correspond to the seven days of the week starting on Monday:

$$f^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 2 & 2 & 2 & 2 \\ 4 & 3 & 3 & 4 & 3 & 3 & 3 \end{bmatrix}$$

The first two states, Low and Medium, only have one possible action, hence the repetition of action zero in the first two rows. The final two rows contain the optimal actions to be taken on each day if the Markov process is in either the High or Very High states. If the MDP is in the High state on Monday, 12 patients should be dismissed and eight on Tuesday through Sunday. If the MDP is in a state of Very High occupancy on Monday or Thursday, then 16 patients are to be dismissed (under the policy f^*) and 12 patients should be dismissed on all other days of the week.

5 Results

Now that the optimal policy has been found, the improvement between the current situation and the optimal policy can be quantified. As the time horizon considered was a calendar week, the weekly steady state distributions of $Q_0, \dots, Q_6$ under the optimal policy f^*will be compared with the corresponding distributions under the existing mode of operations. The comparison of steady state probabilities are shown in Figure 5.1 below. In the figure, each of the four panels corresponds to the four aggregate states Low, Medium, High and Very High. The red curve plots the probabilities of being in the fixed aggregate state on each day of the week under the optimal policy and the blue curve plots the corresponding probabilities under the current mode of operations.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

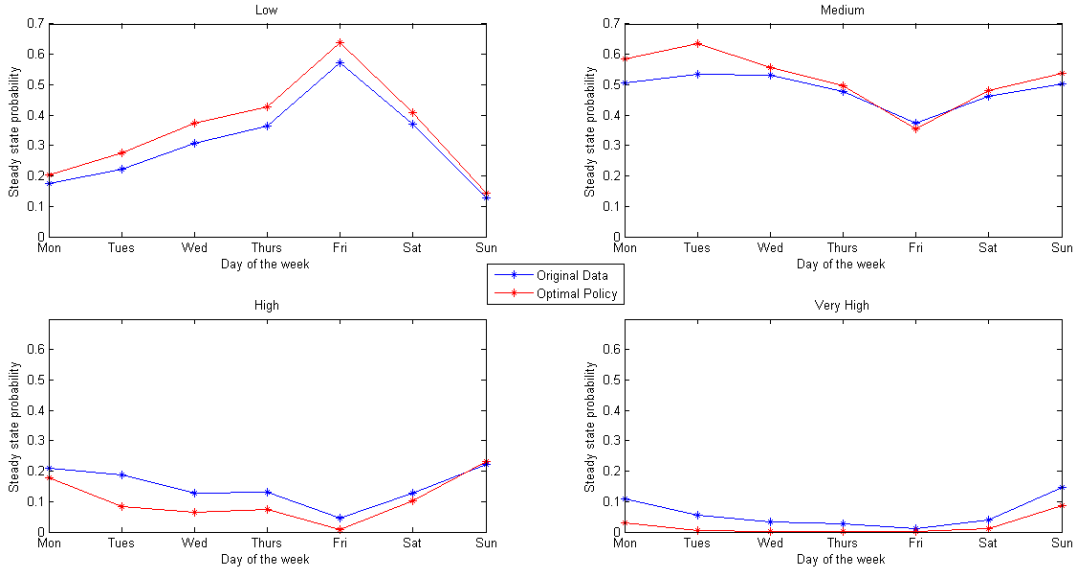

Figure 5.1: Comparison of steady state distributions for weekly probability transition matrices under current conditions and the optimal policy

From Figure 5.1, it can be seen that the optimal policy makes a significant difference in the steady state probabilities for each day of the week. There are increases in the probabilities of being in the two lower occupancy states for nearly every day of the week, demonstrated by the fact that the red curve lies above the blue curve. There are also analogous decreases of the probabilities of being in the upper capacity states for each day of the week, demonstrated by the red curve lying below the blue curve in the bottom two panels. Note the significant improvement in the proportion of days that are considered to have Very High occupancy; it drops to one percent or below for each day between Tuesday and Saturday inclusive. Sunday and Monday are much larger than one percent, but are still lower compared to the current situation by approximately one-third and two-thirds respectively. The decrease in most of the High steady state probabilities is also a desirable outcome.

Figure 5.1 also indicates that there is a slight increase in the High state on a week ending on a Sunday when compared to the current policy and a slight decrease in the Medium state on a Friday. This seems counterintuitive, but those anomalies are offset by corresponding changes in the probabilities of being in the adjacent states.

On Fridays, the decrease in probability of the process being in the Medium state is offset by an increase in being in the Low state. On Sundays, the probability of being in the High state increases slightly compared to the current policy, but the probability of being in the Very High state drops by approximately one-third. This is an advantageous outcome as it is the days of Very High occupancy that are the worst in terms of causing congestion episodes.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

6 Conclusion and Future Improvements

Congestion in hospitals is quickly becoming a considerable hindrance to quality of care and to staff morale. Fortunately, hospital occupancy data exhibit regular patterns, which make it possible to identify trends and subsequently to try to mitigate against congestion. Occupancy levels can be modelled as Markov chains on a week-to-week basis, and can be formulated as Markov Decision Processes with the addition of actions and associated rewards. These decision processes can be used to find an optimal policy that maximises the expected reward over the weekly horizon. Steady state distributions can be used to help quantify the difference between occupancy rates under current policies and under the identified optimal policy. This shows the effectiveness of the optimal policy over the current practice.

The methodology outlined here is generic in the sense that it could be adapted to policies of another hospital. Of course, the parameters would have to be re-estimated and re-calibrated based on the data of that hospital. If hospital managers were interested in the proposed methodology, it is clear that the model could be extended to include more states and more actions so as to capture more accurately the difficulties encountered in real life. Another portion of the model that could be fine-tuned is the definition of the reward functions, which could include a dependency on the state that the process moves to as well as the current state or could be modelled using different base formulae.

7 Acknowledgements

I am indebted to AMSI for the summer scholarship opportunity and to CSIRO for running the Big Day In. I am also grateful to Dr. Mark Mackay from the Flinders University School of Health Sciences for his help in procuring the raw census data and explanations of some of the hospital's operations. My supervisors Prof. Jerzy Filar and Dr. Shaowen Qin provided me with all the help and guidance I needed to complete the project.

8 References

Bellman, R. E. (1957). *Dynamic Programming*. Princeton University Press, Princeton N.J.

Filar, J. A., K. Vrieze (1997). *Competitive Markov Decision Processes*. New York, Springer.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

FitzGerald, G., S. Toloo, J. Rego, J. Ting, P. Aitken, V. Tippett (2012). 'Demand for public hospital emergency department services in Australia: 2000-2001 to 2009-2010.' *Emergency Medicine Australasia* 24(1): 72-78.

Heyman, D.P., M.J. Sobel (1984). *Stochastic Models in Operations Research Vol. II: Stochastic Optimization*. McGraw-Hill, New York.

Howard, R.A (1960). *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, M.A.

Qin, S., J.A. Filar (2014). 'Modelling Hospital Patient flow for Decision Support". To appear in the Proceedings of the 2014 International Conference on Mechanical Engineering and Automation (ICMEA 2014).

SA Health (2014). "Inpatient Dashboard :: SA Health." Retrieved from *http://sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/about+us/our+performance/our+hospital+dashboards/about+the+inpatient+dashboard/inpatient+dashboard* on 28/1/14.

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

Appendix A Daily Transition Matrices

$P_0 =$

0.8293	0.1707	0	0
0.1525	0.7966	0.0508	0
0	0.4694	0.4286	0.102
0	0.0385	0.6538	0.3077

$P_1 =$

0.8654	0.1346	0	0
0.216	0.728	0.056	0
0	0.5	0.4318	0.0682
0	0.3077	0.3077	0.3846

$P_2 =$

0.8333	0.1667	0	0
0.2033	0.6992	0.0976	0
0	0.4333	0.5	0.0667
0	0	0.5	0.5

$P_3 =$

0.8941	0.1059	0	0
0.5135	0.4775	0.009	0
0	0.7742	0.1935	0.0323
0	0.1429	0.5714	0.2857

$P_4 =$

0.6015	0.3985	0	0
0.069	0.6322	0.2759	0.023
0	0	0.5455	0.4545
0	0	0	1

$P_5 =$

0.3488	0.6395	0.0116	0
0	0.5648	0.3704	0.0648
0	0.0333	0.3667	0.6
0	0	0	1

$P_6 =$

0.7333	0.2333	0.0333	0
0.1538	0.7265	0.1111	0.0085
0.0192	0.4808	0.4038	0.0962
0	0.0286	0.4	0.5714

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au

Appendix B Weekly Transition Matrices

$Q_0 =$

0.2101	0.5405	0.1794	0.07
0.1837	0.5182	0.2012	0.0969
0.1539	0.4804	0.2281	0.1377
0.1282	0.4427	0.2521	0.177

$Q_1 =$

0.2625	0.5557	0.1443	0.0375
0.2284	0.5418	0.1784	0.0513
0.1837	0.5108	0.2317	0.0738
0.1459	0.4709	0.2858	0.0974

$Q_2 =$

0.3531	0.5223	0.1017	0.0229
0.3089	0.5319	0.1261	0.0331
0.242	0.5389	0.1678	0.0513
0.152	0.5353	0.2322	0.0806

$Q_3 =$

0.4075	0.466	0.1096	0.0169
0.3652	0.4788	0.1313	0.0248
0.2805	0.4977	0.1787	0.0431
0.1903	0.5026	0.2385	0.0685

$Q_4 =$

0.612	0.3477	0.0327	0.0076
0.5326	0.3974	0.0549	0.015
0.421	0.4589	0.092	0.028
0.3908	0.4745	0.1028	0.0319

$Q_5 =$

0.41	0.4624	0.1034	0.0241
0.3651	0.4655	0.1293	0.0401
0.2992	0.4597	0.1683	0.0728
0.2678	0.4557	0.187	0.0895

$Q_6 =$

0.1558	0.5453	0.2041	0.0949
0.1365	0.5176	0.2191	0.1268
0.1194	0.4866	0.2288	0.1652
0.0934	0.4349	0.2405	0.2313

Postal Address: Building 161
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 8344 6324
Web: www.amsi.org.au