Chapter 2

SOME ANALYTICAL TOOLS

USED IN THE THESIS
2.1 Introduction

In this chapter we describe the analytical tools used in this thesis. They are Markov Decision Processes (MDP), Markov Renewal process (MRP), semi-Markov process (SMP) and semi-Markov decision process (SMDP). The Markov decision processes are simple yet powerful models for sequential decision problems. In these models, we assume that there is a state space; at each time the system occupies a certain state, and the decision maker, or controller, has a set of feasible actions for that state that can be applied. At the next time, the state changes according to some probability distribution which depends only on the current state and action, and does not depend on the past. MDPs are also called controlled Markov chains in the literature, and have a wide range of application areas.

2.2 Markov Decision Processes (MDP)

At the first phase we describe the theory and computational methods developed for MDPs with a finite number of states and actions, and in a discrete-time context, by which we mean that the state transitions occur and the actions are applied at integer times 0, 1, …. Now, we will focus on the probabilistic aspect of our problem formulation: how is the uncertainty modeled?; what is the corresponding mathematical object that we are dealing with?; what is the form of the optimization problems? and what does the Markovian assumption on the state evolution imply?
Model Parameters

The model of an MDP specifies parameters relating to system dynamics and cost:

- a state space $S$;

- for every state a set of feasible actions, which can be jointly represented by the set

  \[ \{(s, U(s)), \forall s \in S\}, \]

  where $U(s)$ is the set of feasible actions at state $s$; we define

  \[ U = \bigcup_{s \in S} U(s) \]

  and call it the action or control space;

- a set of state transition probabilities,

  \[ \{p_{ij}(u), \forall i \in S, u \in U(i)\}; \]

  where $\sum_{j \in S} U(j) = 1$ and

- a per-stage cost function, $c_i(u)$.

Here the state space $S$ and the action space $U(s)$ are assumed to be finite. The economic consequences of the decisions taken at the decision epochs are reflected in receiving a lump sum reward (or pays a cost). This controlled dynamic system is called a discrete-time Markov decision model when the Markov property is satisfied. Note that the one-step costs $c_i(u)$ and the one-step transition probabilities $p_{ij}(u)$ are assumed to be time homogeneous.
Stationary policies

A policy or rule for controlling a system is a prescription for taking actions at each decision epoch. A stationary policy $\pi$ is a rule that always prescribe a single action $\pi_i$ whenever the system is found in state $i$ at a decision epoch.

We define for $n = 0, 1, ...$

$X_n =$ the state of the system at the $n$th decision epoch.

Under he given stationary policy $\pi$, We have

$$P\{X_{n+1} = j / X_{n+1} = i\} = p_{ij}(\pi_i),$$

regardless of past history of the system up to time $n$. Hence under a given stationary policy $\pi$ the stochastic process $\{x_n\}$ is a discrete-time Markov chain with one step transition probabilities $p_{ij}(\pi_i)$. This Markov chain incurs a cost $c_i(\pi_i)$ each time the system visits the state $i$. Thus we can invoke results from Markov chain theory to specify the long-run average cost per unit time under a given stationary policy.

Average cost for a given stationary policy

For a given stationary policy $\pi$, we denote the $n$-step transition probabilities of the corresponding Markov chain $\{x_n\}$ by

$$p^n_{ij}(\pi) = P\{X_n = j / X_0 = i\}, \quad i, j \in S \quad and \quad n = 1, 2, ....$$
where \( p_{ij}^1(\pi) = p_{ij}(\pi) \).

By Chapman-Kolmogorov equations,

\[
p_{ij}^n(\pi) = \sum_{k \in S} p_{ik}^{n-1}(\pi)p_{kj}(\pi_k), \quad n = 2, 3, \ldots
\]

Also we define the expected cost function \( V_n(i, \pi) \) by

\[ V_n(i, \pi) = \text{the total expected costs over the first } n \text{ decision epochs when the initial state is } i \text{ and the policy } \pi \text{ is used.} \]

Thus, we have

\[
V_n(i, \pi) = \sum_{t=0}^{n-1} \sum_{j \in S} p_{ij}^t(\pi)c_j(\pi_j) \quad (2.2.1)
\]

where \( p_{ij}^0(\pi)=1 \) for \( j = i \) and \( p_{ij}^0(\pi)=0 \) for \( j \neq i \). Next we define the average cost function \( g_i(\pi) \) by

\[
g_i(\pi) = \lim_{n \to \infty} \frac{1}{n} V_n(i, \pi), \quad i \in S
\]

The long run average expected cost per unit time is independent of initial state \( i \) when it is assumed that the Markov chain \( \{X_n\} \) corresponding to policy \( \pi \) has no two disjoint closed sets.

In the unichain case we can write

\[
g_i(\pi) = g(\pi), \quad i \in S
\]

Then it follows by the ergodic theorem,

\[
g(\pi) = \sum_{j \in S} c_j(\pi_j)E_j(\pi),
\]
where \( \{E_j(\pi), j \in S\} \) is the unique equilibrium distribution of Markov chain \( \{X_n\} \).

The \( E_j(\pi) \)'s are the unique solution to the system of linear equations

\[
E_j(\pi) = \sum_{k \in S} p_{kj}(\pi_k) E_k(\pi), \quad j \in S
\]

\[
\sum_{j \in S} E_j(\pi) = 1
\]

Moreover, for any \( j \in S \),

\[
E_j(\pi) = \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} p^n_{ij}(\pi) \quad \forall \quad i \in S
\] (2.2.2)

We have that \( g(\pi) \) is the expected value. Also with probability 1, the long-run actual average cost per unit time = \( g(\pi) \) independently of the initial state.

**Average cost optimal policy**

The optimization problem is now to find a policy with the minimum cost, with respect to the chosen cost criterion, for a given or every initial state.

A stationary policy \( \pi^* \) is said to be average cost optimal if

\[
g_i(\pi^*) \leq g_i(\pi)
\]

for each stationary policy \( \pi \) uniformly in the state \( i \). It is stated without proof that an average cost minimal policy \( \pi^* \) always exists. Moreover, policy \( \pi^* \) is not only optimal among the class of stationary policies but it is also optimal among the class of all conceivable policies (see ref: Derman(1970)).

68
2.3 Markov Renewal Process and Semi-Markov Processes

Some systems follow a Markov property not at all point of time but only for a special increasing stopping times. These times are the state changes times of the considered stochastic process.

we present below basic results and definitions of a Markov renewal process.

Consider an at most countable set $S$ say, a two-dimensional stochastic process $(X, T) = (X_n, T_n, n \in \mathbb{N})$ where the random variable (r.v.) $X_n$ takes values in $S$ and the r.v. $T_n$ takes values in $\mathbb{R}_+$ and satisfies $0 = S_0 \leq S_1 \leq S_2 \leq \ldots$

**Definition 2.3.1** The stochastic process $(X, T)$ is called a Markov Renewal Process (MRP) if it satisfies the following relation

$$P[X_{n+1} = j, T_{n+1} - T_n \leq t | X_0, \ldots, X_{n-1}, X_n = i; T_0, \ldots, T_n] = P[X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i] = Q_{i,j}(t)$$

for all $n \in \mathbb{N}, j \in S$ and $t \in \mathbb{R}_+$

The set $S$ is called state space of the MRP the function $Q_{i,j}(t)$ form semi-Markov kernel.

From these relations it is clear that $(X_n)$ is a Markov chain with state space $S$. 
and transition probability $P(i, j) = Q_{i,j}(\infty)$. It is called embedded Markov chain. For every $i \in S$, we have $P(i, i) = 0$.

Now, we define, the counting process $\{N(t), t \geq 0\}$ associated to the point process $\{T_n, n \geq 0\}$ i.e. for each time $t \geq 0$ the r.v $N(t)$ is

$$N(t) := \sup\{n : T_n \leq t\}$$

and define continuous time process $Z = \{Z(t), t \in \mathbb{R}\}$ by

$$Z(t) := X_{N(t)}.$$

Then the process $Z$ is called semi-Markov process (SMP). Also define

$$P_{i,j}(t) = P[Z(t) = j | Z(t) = 0]$$

$$H_i(t) = \sum_{j \in S} Q_{i,j}(t),$$

$$m_i = \int_0^\infty [1 - H_i(u)]du.$$

We have the following particular classes of the MRP

(1) Discrete time Markov chain

$$Q_{i,j}(t) = P(i, j) I\{t \geq 0\}, \quad \text{for all } i, j \in S, \text{ and } t \geq 0.$$  

(2) Continuous time Markov chain

$$Q_{i,j}(t) = P(i, j) (1 - e^{-\lambda_j t}) \quad \text{for all } i, j \in S, \text{ and } t \geq 0.$$
(3) Renewal Process :

(a) Ordinary: It is an MRP with two states $S = \{0, 1\}$,

$$P(0,1)=P(1,0)=1 \text{ and } Q_{01}(\cdot) = F(\cdot),$$
where $F$ is the common distribution function of the inter-arrival times of the renewal process.

(4) Modified or delayed: It is an MRP with three states $S = \{0, 1, 2\}$,

$$P(0,1)=1, P(1,2)=P(2,1)=1 \text{ and 0 elsewhere and } Q_{01}(\cdot) = F_0(\cdot), \quad Q_{12}(\cdot) = Q_{21}(\cdot) = F(\cdot),$$
where $F_0$ is the common distribution function of the first arrival time and $F$ is the common distribution function of the inter-arrival times of the renewal process.

(5) Alternating: It is an MRP with two states $S = \{0, 1\}$,

$$P(0,1)=P(1,0)=1 \text{ and 0 elsewhere and } Q_{01}(\cdot) = F(\cdot), \quad Q_{10}(\cdot) = G(\cdot)$$
where $F$ and $G$ are the common distribution function corresponding to the odd and even inter-arrival times.

For a systematic account of these topics, covering the Markov renewal equation and its solution one may refer to Cinlar(1975). As an extension of the basic classical limit theorems in probability theory to the semi-Markov setting those are available in the literature, some limit theorems useful for reliability/survival analysis.

As noted above, in the SMP environment, two random variables run simultaneously.

$$X_n : \Omega \rightarrow S, \quad T_n : \Omega \rightarrow R, n \in N.$$ 

$X_n$ with state space, say, $S = \{S_1, \ldots, S_m\}$ represents the state at the $n$th transition.
In the health care environment, the elements of $S$ represent all the possible stages in which the disease may show level of seriousness. $T_n$, with state space equal to $\mathbb{R}$, represents the time of the $n$th transition. In this way, we can not only consider the randomness of the states but also the randomness of the time elapsed in each state. The process $(X_n, T_n)$ is assumed to be a homogeneous Markovian renewal process.

Furthermore, it is necessary to introduce the probability that the process will leave state $i$ in a time $t$ as

$$H_i(t) = P[T_{n+1} - T_n \leq t | X_n = i].$$

Obviously,

$$H_i(t) = \sum_{j=1}^{m} Q_{i,j}(t).$$

It is now possible to define the distribution function of the waiting time in each state $i$, given that the state successively occupied is known,

$$G_{i,j}(t) = P[T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j]$$

Obviously, the related probabilities can be obtained by means of the following formula:

$$G_{ij}(t) = \begin{cases} \frac{Q_{i,j}(t)}{P(i,j)}, & \text{if } P(i,j) = 0 \\ 1, & \text{if } P(i,j) \neq 0 \end{cases}$$

The main difference between a continuous time Markov process and a semi-Markov process lies in the distribution functions $G_{ij}(t)$. In a Markov environment this function must be a negative exponential function. On the other hand, in the semi-Markov case, the distribution functions $G_{ij}(t)$ can be of any type. This means that the transition intensity can be decreasing or increasing.
If we apply the semi-Markov model in the health care environment, we can consider, by means of the $G_{ij}(t)$, the problem given by the duration of the time spent inside one of the possible disease states.

Now the homogeneous SMP, $Z = \{Z(t), t \in \mathbb{R}\}$ represents, for each waiting time, the state occupied by the process

$$Z(t) = X_{N(t)}, \quad \text{where} \quad N_t = \text{sup}\{n : T_n \leq t \}.$$ 

The transition probabilities are defined in the following way:

$$\phi_{ij}(t) = P[Z(t) = j | Z(0) = i]$$

They are obtained by solving the following evolution equations:

$$\phi_{ij}(t) = \delta_{i,j}(1 - H_i(t)) + \sum_{\beta=1}^{m} \int_{0}^{t} Q'_{i\beta}(\vartheta) \phi_{\beta j}(t - \vartheta) d\vartheta, \quad (2.3.1)$$

where $\delta_{i,j}$ represents the Kronecker delta.

The first addendum of formula (2.3.1) gives the probability that the system does not undergo transitions up to time $t$ given that it was in state $i$ at an initial time 0. In predicting the disease evolution model, it represents the probability that the infected patient does not shift to any new stage in a time $t$. In the second addendum, $Q'_{i\beta}(\vartheta)$ is the derivative at a time $\vartheta$ of $Q_{i,\beta}(\vartheta)$ and it represents the probability that the system remained in a state $i$ up to the time $\vartheta$ and that it shifted to state $\beta$ exactly at a time $\vartheta$. After the transition, the system will shift to state $j$ following one of all the possible trajectories from state $\beta$ to state $j$ within a time $t - \vartheta$. In disease evolution model, it means that up to a time an infected subject remains in the state $i$. At the
time $\vartheta$, the patient moves into a new stage $\beta$ and then reaches state $j$ following one of the possible trajectories in some time $t - \vartheta$.

### 2.4 Semi-Markov Decision Processes (SMDPs)

In Markov decision processes (MDPs) the decisions are taken at each of a sequence of unit time intervals. The Semi-Markov decision processes (SMDPs) generalize MDPs by allowing the decision maker to choose actions whenever the system state changes, modeling the system evolution in continuous time and allowing the sojourn time in a particular state to follow an arbitrary probability distribution. The system state may change several times between decision epochs; only the state at a decision epoch is relevant to the decision maker. If transition times between states are distributed exponentially, we refer to the process as continuous-time Markov decision process (CTMDP).

The model of an SMDP specifies parameters relating to system dynamics and cost are specified similarly as in MDP:

- At a decision epoch the system occupies a state $s \in S$, $S$ is the state space;

- for every state $s$ a set of feasible actions called action or control space $U(s)$, which can be jointly represented by the set

\[\{(s, U(s)), \forall s \in S\},\]
Also we define

\[ U = \bigcup_{s \in S} U(s); \]

- a set of state transition probabilities,

\[ \{Q_{ij}(t|u), \forall s \in S, u \in U(s)\}; \]

In most applications \( Q_{ij}(t|u) \) is not provided directly, but instead \( F_i(t|u) \) and \( p_{ij}(t|u) \) are used. The \( F_i(t|u) \) denotes the probability that the next decision epoch occurs within \( t \) time units, given that action \( u \in U \) is chosen in state \( i \). The quantity \( p_{ij}(t|u) \) denotes the probability that the system occupies state \( j \) in \( t \) time units after the decision epoch given \( i \) and \( u \). (If the natural process does not change state until the next decision epoch, \( p_{ij}(t|u) = 1 \) for all \( t \).

The following assumption is needed to guarantee that there will not be an infinite number of decision epochs within finite time:

There exists \( \epsilon > 0 \) and \( \delta > 0 \) such that \( F_i(t|u) \leq 1 - \epsilon \) for all \( u \in U \) and \( i \in S \)

Two special cases for \( F_i(t|u) \):

- When \( F_i(t|u) = 1 - e^{-\beta(i,u)t} \)
  we refer to this as a continuous-time Markov decision process.

- When \( F_i(t|u) = 0 \) if \( t \leq t' \); \( =1 \) if \( t > t' \)
  for some fixed \( t \) for all \( i \) and \( u \), we obtain a discrete-time MDP.
When decision maker chooses action $u$ in state $i$, he receives a lump sum reward (or pays a cost) $g(i, u)$. In addition to that, he accrues a reward (or incurs a cost) at rate $c(j', i, u)$ as long as the natural process occupies state $j$, and action $u$ was chosen in state $i$ at preceding decision epoch.

**Decision rules and policies**

In SMDP, the decision rules may be deterministic or randomized, Markovian or history dependent. Let us define a policy $\pi = (d_1, d_2...)$ decision rule $d_1$ is used at $t_0 = 0$.

In discounted model for a policy $\pi \in \Pi$, let us denote $v^\pi_\alpha(s)$ as the expected infinite-horizon discounted reward with the discount factor $\alpha$, given that the process occupies state $s$ in the first decision epoch:

$$v^\pi_\alpha(s) \equiv E_s \left( \sum_{n=0}^{\infty} e^{-\alpha T_n} \left[ g(X_n, Y_n) + \int_{T_n}^{T_{n+1}} e^{-\alpha(t-T_n)} c(w_t, X_n, Y_n) dt \right] \right)$$

where $T_0, T_1, ...$ represent the times of successive decision epochs, $w_t$ denotes the state of the natural process at time $t$. Define the value of a discounted SMDP by

$$v^*_\alpha(s) \equiv \sup_{\pi \in \Pi} v^\pi_\alpha(s)$$

The goal is to find a policy $\pi^*$ for which $v^{{\pi}^*}_\alpha(s) = v^*_\alpha(s)$.

In other words, the objectives in these problems, is to maximize the expected...
discounted reward or minimize the expected discounted total costs and obtain the best policy. We confine ourselves to the optimality criterion of the total expected discounted cost. This criterion is found to be more appropriate in many applications, particularly in biomedical and reliability studies. An alternative criterion is the long run average cost per unit time which is more appropriate when many state transitions occur within a relatively short time just as in the stochastic control problems in telecommunication application.