10th workshop on Markov processes and related topics

Finite-horizon Optimization for continuous-time Markov decision processes

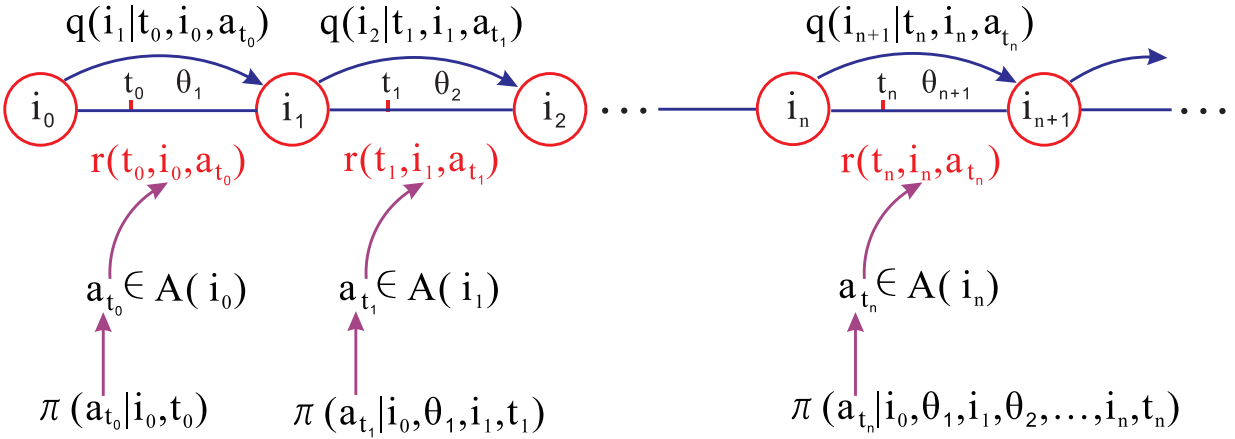
Xianping Guo (with X.X. Huang, and Y.H. Huang)
Sun Yat-Sen University
Email: mcsgxp@mail.sysu.edu.cn

14-18 July 2014, Xian

Outline

- The model of continuous-time MDP
- The optimality problem
- Extensions of some results for "MC"
- Existence of optimal policies
- An example

1. The model of continuous-time MDP



- i_n: State at time t_n, which is in the state space S
- θ_{n+1}: The holding time at state i_n
- $q(j|t,i,a)$: Transition rates depending on action $a \in A(i)$
- $r(t,i,a)$: Reward function of time t, states i and actions a

Let

- $\Omega^0 := (S \times (0, \infty))^\infty$,
- $\Omega := \Omega^0 \bigcup \{(i_0, \theta_1, i_1, \ldots, i_k, \infty, \Delta, \infty, \ldots) |\ i_0 \in S, i_l \in S, \theta_l \in (0, \infty)$, for each $1 \le l \le k, k \ge 1\}$, with $\Delta \notin S$
- $\mathcal{F}$: σ-algebra on Ω. For $e = (i_0, \theta_1, i_1, \ldots, \theta_k, i_k, \cdots) \in \Omega$, let $T_k(e) := \theta_1 + \theta_2 + \ldots + \theta_k, T_\infty(e) := \lim_{k \to \infty} T_k(e)$
- Define the state process $\{x_t, t \ge 0\}$ by

$$x_t := \sum_{k \ge 0} I_{\{T_k \le t < T_{k+1}\}} i_k + \Delta I_{\{t \ge T_\infty\}}, \qquad \text{for } t \ge 0. \quad (1)$$

Here and below, I_E stands for the indicator function on E.

- Randomized history-dependent policies $\pi(da|e,t)$: is defined by the following expression

$$\pi(da|e,t) = \sum_{k\geq 0} I_{\{T_k<t\leq T_{k+1}\}}\pi^k(da|i_0,\theta_1,i_1,\ldots,\theta_k,i_k,t-T_k) + I_{\{t=0\}}\pi^0(da|i_0,0) + I_{\{t\geq T_\infty\}}\delta_{a_\Delta}(da), \quad (2)$$

  depending on histories $(i_0,\theta_1,i_1,\ldots,\theta_k,i_k)$

- Markov policies π: $\pi(da|i,t)$
- Π^m: The class of all Markov policies.
- Π : The class of all randomized history-dependent policies.

2. The optimality problem

Given the transition rates $q(j|i,a)(i,j \in S, a \in A(i))$, each $\pi \in \Pi$ (with a fixed $i \in S$) ensures a unique p.m. P_i^π on $\mathcal{F}$.

$$V_\pi(0,i) := \mathbb{E}_i^\pi \left[\int_0^T \int_A r(s,x_s,a)\pi(da|e,s)ds + g(T,x_T) \right]$$

$$V_\pi(t,i) := \mathbb{E}_{t,i}^\pi \left[\int_t^T r(s,x_s,a)\pi(da|x_s,s)ds + g(T,x_T) \right]$$

for Markov policy π, where $\mathbb{E}_{t,i}^\pi[X] := \mathbb{E}_i^\pi[X|x_t = i]$.

The value function $V^*(t,i)$ $(i \in S, t \geq 0)$ is defined by

$$V^*(0,i) = \sup_{\pi \in \Pi} V_\pi(0,i), \qquad V^*(t,i) := \sup_{\pi \in \Pi^m} V_\pi(t,i), \quad t > 0.$$

Optimal policy π^*: $V(0,\pi^*) \geq V^*(0,i)$ for all $i \in S$.

Remark 1:

Concerning the value function $V^*(t,i)$, we state the unsolved problems by Yushkevich (Theory Probab. Appl., 22, 215–235, 1977): "Unsolved problems. In analogy to the discrete time case it would be desirable to extend Theorems 4.1 and 4.2 to arbitrary summable models and in Theorems 5.1 and 5.2 to do away with the required boundedness of v_t".

In the unsolved problems, which are also called open questions by H.-J. Engelbert in [MR0458603 (56#16803)], v_t is the value function here.

3. Extensions of some results for "MC"

To guarantee the regularity of $\{x_t\}$ and the finiteness of $V^*(t,i)$, we give the following condition:

Assumption A. A function $w \geq 1$, positive constants $c, b \geq$, subsets $S_k \uparrow S$, such that

(1) $\sum_{j\in S} w(j)q(j|t,i,a) \leq cw(i) + b, i \in S, a \in A(i)$;

(2) $\inf_{i\notin S_k} w(i) \uparrow +\infty$ as $k \to \infty$, with $\inf \emptyset := \infty$;

(3) $\sup_{a\in A(i), i\in S_k, t\geq 0} |q(i|t,i,a)| < \infty$ for $k \geq 1$;

(4) $|r(t,i,a)| + |g(T,i)| \leq Mw(i)$, with some $M > 0$.

Theorem 1. Under Assumptions A(1)-A(3), we have

(a) $P_i^\pi(T_\infty = \infty) \equiv 1$, and $P_i^\pi(x_t \in S) \equiv 1$;

(b) the analog of the forward Kolmogorov equation holds:

$$P_i^\pi(x_t = j) = \delta_{ij} + E_i^\pi \left[\int_0^t \int_A \pi(da|e, s) q(j|s, x_s, a) ds \right];$$

(c) if the additional Assumption A(4) holds, then

$$(c_1) \quad |V_\pi(0, i)| \leq (T+1) M_1 e^{cT} [\omega(i) + \tfrac{b}{c}], \qquad \pi \in \Pi;$$

$$(c_2) \quad |V_\pi(t, i)| \leq (T+1) M_1 e^{c(T-t)} [\omega(i) + \tfrac{b}{c}], \qquad \pi \in \Pi^m.$$

Remark 2: Theorem 1 (b) gives the analog of the forward Kolmogorov equation.

To further derive the analog of Ito-Dynkin formula for the process $\{x(t)\}$, we consider the condition below:

Assumption B. With ω as in Assumption A, a function $w' \geq 1$, constants $c' > 0$, $b' \geq 0$ and $M_2 > 0$ such that

$$q^*(i)w(i) \leq M_2 w'(i), \text{ and } \sum_{j \in S} w'(j) q(j|t,i,a) \leq c' w'(i) + b'$$

where $q^*(i) := \sup_{t \geq 0, a \in A(i)} |q(i|t,i,a)|$.

$C^{1,0}_{\omega,\omega'} := \{\varphi : \varphi(t,i)$ has the derivative $\varphi'(t,i)$ at a.e. t, and

$$\sup_{i,t} \frac{|\varphi(t,i)}{w(i)} < \infty, \sup_{i,t} \frac{|\varphi'(t,i)}{w(i) + w'(i)} < \infty\}$$

Theorem 2. Under Assumptions A and B, for each $\varphi \in C^{1,0}_{\omega,\omega'}$, the following assertions hold.

(a) (The Ito-Dynkin formula): For each $\pi \in \Pi^m$,

$$\mathbb{E}^{\pi}_{t,i}\left[\int_t^T \left(\varphi'(s,x_s) + \sum_{j\in S} \varphi(s,j) q(j|s,x_s,a)\pi(da|x_s,s)\right) ds \right.$$
$$= \mathbb{E}^{\pi}_{t,i}\varphi(T,x_T) - \varphi(t,i).$$

(b) (The analog of Ito-Dynkin formula): For every $\pi \in \Pi$,

$$\mathbb{E}^{\pi}_{0,i}\left[\int_0^T \left(\varphi'(s,x_s) + \sum_{j\in S} \int_A \varphi(s,j) q(j|s,x_s,a)\pi(da|e,s)\right) \right.$$
$$= \mathbb{E}^{\pi}_{0,i}\varphi(T,x_T) - \varphi(0,i).$$

Theorem 3. Under Assumptions A and B, for $\pi \in \Pi^m$, $V_\pi(t,i)$ is a unique solution in $C^{1,0}_{\omega,\omega'}$ of the following equation

$$\begin{cases} \varphi'(t,i) + r(t,i,\pi_t) + \sum_{j\in S} \varphi(t,j) q(j|t,i,\pi_t) = 0 \\ \varphi(T,i) = g(T,i) \end{cases}$$

where, $u(s,i,\pi_t) := \int_{A(i)} u(t,i,a)\pi(da|i,t)$.

Theorem 4. Under Assumptions A and B, if there exists $\varphi \in C^{1,0}_{\omega,\omega'}$, such that, for all $t \geq 0, i \in S, a \in A(i)$,

$$\begin{cases} \varphi'(t,i) + r(t,i,a) + \sum_{j \in S} \varphi(t,j) q(j|t,i,a) \leq 0 \\ \varphi(T,i) = g(T,i) \end{cases}$$

then,

(a) $V_\pi(0,i) \leq \varphi(0,i)$, for all $\pi \in \Pi, i \in S$;

(b) $V_\pi(t,i) \leq \varphi(t,i)$, for all $\pi \in \Pi^m, t \geq 0, i \in S$.

Remark 3: A key point is how to establish the existence of a function φ in Theorem 4.

4. Existence of optimal Markov policies

The arguments are given by an approximation technique.

Lemma 5. Suppose that the transition rates are bounded (i.e., $\sup_{i\in S,a\in A(i)} |q(i|i,a)| < \infty$) and Assumption A is satisfied. Then, the following assertions hold.

(a) There exists a unique $\varphi(t,i)$ in $C^{1,0}_{\omega,\omega}$ satisfying the following *optimality equation* (OE):

$$\begin{cases} \phi'(t,i) + \sup\limits_{a\in A(i)} [r(t,i,a) + \sum\limits_{j\in S} \phi(t,j) q(j|t,i,a)] = 0, \\ \phi(T,i) = g(T,i), \end{cases}$$

(b) $\varphi(t,i) = V^*(t,i)$, with $\varphi(t,i)$ as in (a) above.

Proof. (i). Show that the OE is equivalent to the equation:

$$\begin{aligned}\psi(t,i) &= e^{\beta t}g(T,i) \\ &\quad + e^{\beta t}\int_t^T \sup_{a\in A(i)}\left[r(s,i,a) + e^{-\beta s}\sum_{j\in S}\psi(s,j)q(j|s,i,a)\right]ds.\end{aligned}$$

(ii). Define an operator G by

$$\begin{aligned}G\psi(t,i) &= e^{\beta t}g(T,i) \\ &\quad + e^{\beta t}\int_t^T \sup_{a\in A(i)}\left[r(s,i,a) + e^{-\beta s}\sum_{j\in S}\psi(s,j)q(j|s,i,a)\right]d\end{aligned}$$

Since the transition rates are bounded, we can prove that G is a contraction operator on a Banach space.

Let $\psi^*(t,i)$ be the fixed point of G, $\varphi(t,i) := e^{-\beta t}\psi^*(t,i)$. Then, $\varphi(t,i)$ is a required solution for part (a).

(iii). By Theorem 4 and the generalization of the measurable selection theorem, we can prove (b).

The condition is given for the existence of an optimal policy.

Assumption C (Continuity-compactness conditions):

(i) $A(i)$ is compact for every $i \in S$;

(ii) For each $i, j \in S, t \geq 0$, the functions $q(j|t,i,a), r(t,i,a)$, and $\sum_{j\in S} \omega(j) q(j|t,i,a)$ are continuous in $a \in A(i)$.

Theorem 6. Under Assumptions A, B and C, the following assertions hold.

(a) There exists a unique $\varphi(t,i)$ in $C^{1,0}_{\omega,\omega'}$ satisfying the *optimality equation*.

(b) $\varphi(t,i) = V^*(t,i)$ for all (t,i), with $\varphi(t,i)$ as in (a) above.

(c) There exists a Markov policy $f^* \in \Pi^m$ such that

$$\varphi'(t,i) + r(t,i,f^*(t,i)) + \sum_{j\in S} \varphi(t,j) q(j|t,i,f^*(t,i)) = 0$$

and the Markov policy f^* is optimal.

Proof of Part (a)–(i). Without loss of generalization, let $S := \{0, 1, \ldots, n, \ldots\}$, and $S_n := \{0, 1, \ldots, n\}$, and

$$q_n(j|t, i, a) := \begin{cases} q(j|t, i, a), & i \in S_n, a \in A(i), \\ 0, & \text{otherwise.} \end{cases}$$

Thus, obtain a sequence of models $\{\mathcal{M}_n\}$ of CTMDPs:

$$\mathcal{M}_n := \{S, A, (A(i), i \in S), r(t, i, a), q_n(j|t, i, a)\} \qquad (3)$$

Obviously, for each model $\mathcal{M}_n$, the Assumptions A and B still hold, and the corresponding transition rates are bounded.

Proof of Part (a)–(ii). Let $u_n(t,i)$ be the unique solution to the OE for the $\mathcal{M}_n$. Thus,

$$u_n(t,i) = g(T,i) + \int_t^T \sup_{a\in A(i)} [r(s,i,a) + \sum_{j\in S} u_n(s,j) q_n(j|s,i,a)] ds$$

We can prove that $\{u_n(s,i), n \geq 1\}$ is equicontinuous in (s,i). Ascoli theorem ensures the existence of a subsequence $\{u_{n_k}(t,i), k \geq 1\}$ of $\{u_n(t,i), n \geq 1\}$ and a continuous function φ such that

$$\lim_{k\to\infty} u_{n_k}(t,i) = \varphi(t,i), \text{ and } |\varphi(t,i)| \leq D\omega(i)$$

Proof of Part (a)–(iii).

We further show $\lim_{k\to\infty} H_{n_k}(s,i) = H(s,i)$, where

$$H_n(s,i) := \sup_{a\in A(i)} [r(s,i,a) + \sum_{j\in S} u_n(s,j) q_n(j|s,i,a)];$$

$$H(s,i) := \sup_{a\in A(i)} [r(s,i,a) + \sum_{j\in S} \varphi(s,j) q(j|s,i,a)].$$

Hence,

$$\begin{aligned}\varphi(t,i) \;=\; & g(T,i) \\ & + \int_t^T \sup_{a\in A(i)} [r(s,i,a) + \sum_{j\in S} \varphi(s,j) q(j|s,i,a)] ds,\end{aligned}$$

which is equivalent to the OE.

Proof of Parts (b)-(c): Using the measurable selection theorem, by Part (a) and Theorem 4 we see that parts (b) and (c) are true.

5. An example

A controlled birth-death system: For $a := (a_1, a_2)$

$$q(1|t, 0, a) = -q(0|t, 0, a) := \lambda(t) + a_1 \tag{4}$$

where a_1 is explained as an immigration parameter.

$$q(j|t, i, a) := \begin{cases} \lambda(t)i + a_1 & \text{if } j = i + 1, \\ -[\lambda(t) + \mu(t)]i - a_1 - a_2 & \text{if } j = i, \\ \mu(t)i + a_2 & \text{if } j = i - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

for each $i \geq 1, a = (a_1, a_2) \in A(i)$.

$\mathbf{C_1}$. $\lambda(t)$ and $\mu(t)$: continuous, non-negative, and bounded.

$\mathbf{C_2}$. $A(i) := [-\lambda_1 i, (1+\lambda_2)(1+i)] \times [-\mu_1 i, (1+\mu_2)(1+i)]$

where $\delta_1 := \inf_{t\geq 0} \delta(t), \delta_2 := \sup_{t\geq 0} \delta(t)$, for $\delta \in \{\lambda, \mu\}$.

$\mathbf{C_3}$. $r(t,i,a)$ is continuous in a; and

$$|g(T,i)| \leq M(i^n+1), \quad |r(t,i,a)| \leq M(i^n+1),$$

where $n \geq 1$ is some integer.

Proposition 7.

(a) Under $\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$, the controlled birth-death system satisfies Assumptions A, B, and C. Therefore (by Theorem 6), there exists an optimal Markov policy.

(b) (**A special case**): Suppose that, in addition,

$$\lambda(t) = \mu(t) \equiv 0,\ g(t,i) = 0, \quad A(i) = [0,i] \times [0,2i];$$

$r(t,i,a_1,a_2) = -2i + (T+3-3e^{t-\frac{T}{2}})a_1 + (\frac{3}{2}e^{t-\frac{T}{2}} - \frac{3}{2} - T)a_2$ for $t \in [0, \frac{T}{2})$, and $r(t,i,a_1,a_2) = -2i + (\frac{5T}{2} - 3t)a_1 + (t - \frac{3T}{2})a_2$ for $t \in [\frac{T}{2}, T]$, where $(a_1,a_2) \in A(i)$. Then,

$$V^*(t,i) = \begin{cases} -i(2+T-2e^{t-\frac{T}{2}}),\ i \geq 0,\ t \in [0, \dfrac{T}{2}), \\ -2i(T-t),\ i \geq 0,\ t \in [\frac{T}{2}, T]; \end{cases}$$

which is unbounded in i, and

$$f^*(t,i) = \begin{cases} (i, 2i), & t \in [0, \dfrac{T}{2}), \\ (0,0), & t \in [\frac{T}{2}, T]. \end{cases}$$

Many Thanks !!!