Operations Research Letters 44 (2016) 457–462

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Continuous-time Markov decision processes under the risk-sensitive average cost criterion

Qingda Wei [a], Xian Chen [b,*]

[a] School of Economics and Finance, Huaqiao University, Quanzhou, 362021, PR China

[b] School of Mathematical Sciences, Peking University, Beijing, 100871, PR China

ARTICLE INFO

Article history:
Received 2 November 2015
Accepted 25 April 2016
Available online 4 May 2016

Keywords:
Continuous-time Markov decision processes
Risk-sensitive average cost criterion
Optimality equation
Optimal policy

ABSTRACT

This paper studies continuous-time Markov decision processes under the risk-sensitive average cost criterion. The state space is a finite set, the action space is a Borel space, the cost and transition rates are bounded, and the risk-sensitivity coefficient can take any positive real number. Under the mild conditions, we develop a new approach to establish the existence of a solution to the risk-sensitive average cost optimality equation and obtain the existence of an optimal deterministic stationary policy.

1. Introduction

Continuous-time Markov decision processes (CTMDPs) have wide applications, such as the queueing systems, control of the epidemic, telecommunication, population processes, inventory management; see, for instance, [6,8,9]. The expected average cost criterion is a commonly used optimality criterion in the theory of CTMDPs and has been widely studied under the different sets of optimality conditions; see, for instance, [6,9,10] and the references therein. The random costs incurred during the finite time interval are evaluated by the mathematical expectation in the definition of the expected average cost criterion. In other words, the expected average cost criterion assumes that the decision-makers are risk-neutral. However, different decision-makers may have different risk preferences in the real-world applications. Hence, it is necessary for us to consider the attitude of a decision-maker towards the risk in the definition of the average cost criterion. As is well known, the utility function is an important tool to characterize the risk preferences of the decision-makers. In particular, the exponential utility function is a commonly used utility function and has been applied to reflect the risk attitudes of the decision-makers towards the random costs incurred in the MDPs; see, for instance, [2–4,7] for discrete-time MDPs and [5] for CTMDPs. The average optimality criterion in [2–5,7] is called risk-sensitive average cost criterion because the risk preferences of the decision-makers are taken into consideration. To the best of our knowledge, [5] is the first work to study the risk-sensitive average cost criterion for CTMDPs. The state space is a denumerable set, the cost rate function is nonnegative and bounded, the transition rates are bounded and satisfy the irreducibility condition and some Lyapunov-like inequality, and the risk-sensitivity coefficient of the exponential utility function is positive and satisfies some additional relation in [5].

In this paper we further study the risk-sensitive average cost criterion in the class of all randomized Markov policies for CTMDPs. The state space is a finite set and the action space is a Borel space. The cost rate function is bounded and allowed to take both nonnegative and negative values. The transition rates are bounded and the risk-sensitivity coefficient is allowed to take any positive real number. Under the irreducibility condition and the continuity and compactness conditions, we employ a new approach to establish the existence of a solution to the risk-sensitive average cost optimality equation, from which the existence of optimal policies is shown. More precisely, we first introduce an auxiliary risk-sensitive first passage optimization problem and obtain the properties of the optimal value function of the risk-sensitive first passage problem (see Theorem 3.1). Then using the Feynman–Kac formula and the results on the risk-sensitive first passage optimization problem, we show that the pair of the optimal value functions of the risk-sensitive average cost criterion and the risk-sensitive first passage problem is a

* Corresponding author.
E-mail addresses: weiqd@hqu.edu.cn (Q. Wei), chenxian@amss.ac.cn (X. Chen).

http://dx.doi.org/10.1016/j.orl.2016.04.010

solution to the risk-sensitive average cost optimality equation and that there exists an optimal deterministic stationary policy in the class of all randomized Markov policies (see Theorem 3.2). As far as we can tell, the risk-sensitive first passage optimization problem for CTMDPs is discussed for the first time in this paper. Moreover, since we remove the nonnegativity of the cost rate function, the Lyapunov-like inequality imposed on the transition rates and the additional relation required for the positive risk-sensitivity coefficient in [5], the optimality conditions in this paper are weaker than those in [5] except that the state space is a finite set. Furthermore, we deal with the risk-sensitive average cost criterion in a more general class of policies than that in [5] which investigates this criterion in the class of all deterministic stationary policies.

The rest of this paper is organized as follows. In Section 2, we introduce the decision model and the risk-sensitive average cost criterion. In Section 3, we give the optimality conditions and the main results whose proofs are presented in Section 4.

2. The decision model

The decision model we are concerned with is composed of the following components

$$\{S, A, (A(i),\ i \in S), q(j|i,a), c(i,a)\},$$

where the state space S is a finite set endowed with the discrete topology, the action space A is a Borel space with the Borel σ-algebra $\mathcal{B}(A)$, and $A(i) \in \mathcal{B}(A)$ is the set of all admissible actions in state $i \in S$. Let $K := \{(i,a)|i \in S, a \in A(i)\}$ be the set of all admissible state-action pairs. The real-valued transition rate $q(j|i,a)$ satisfies the following properties: (i) For each fixed $i, j \in S$, $q(j|i,a)$ is measurable in $a \in A(i)$; (ii) $q(j|i,a) \geq 0$ for all $(i,a) \in K$ and $j \neq i$; (iii) $\sum_{j\in S} q(j|i,a) = 0$ for all $(i,a) \in K$. The real-valued cost rate function $c(i,a)$ is measurable in $a \in A(i)$ for each $i \in S$.

A continuous-time Markov decision process evolves as follows. A decision-maker observes continuously the state of a dynamical system. When the system is in state $i \in S$, an action $a \in A(i)$ is chosen by the decision-maker according to some decision rule and such an intervention has the following consequences: (i) a cost is incurred at the rate $c(i,a)$; (ii) the system remains in the state i for a random time following the exponential distribution with the tail function given by $e^{q(i|i,a)t}$, and then jumps to a new state $j \neq i$ with the probability $-\frac{q(j|i,a)}{q(i|i,a)}$ (we make a convention that $\frac{0}{0} := 0$).

Let $S_\infty := S \cup \{i_\infty\}$ with an isolated point $i_\infty \notin S$, $\mathbb{R}_+ := (0, +\infty)$, $\Omega^0 := (S \times \mathbb{R}_+)^\infty$, $\Omega := \Omega^0 \cup \{(i_0, \theta_1, i_1, \ldots, \theta_{m-1}, i_{m-1}, \infty, i_\infty, \infty, i_\infty, \ldots)|i_0 \in S,\ i_l \in S,\ \theta_l \in \mathbb{R}_+$ for each $1 \leq l \leq m-1,\ m \geq 2\}$, and $\mathcal{F}$ be the Borel σ-algebra of Ω. For each $\omega = (i_0, \theta_1, i_1, \ldots) \in \Omega$, define $X_0(\omega) := i_0, T_0(\omega) := 0, X_m(\omega) := i_m, T_m(\omega) := \theta_1 + \theta_2 + \cdots + \theta_m$ for $m \geq 1$, $T_\infty(\omega) := \lim_{m\to\infty} T_m(\omega)$, and the state process

$$\xi_t(\omega) := \sum_{m\geq 0} I_{\{T_m \leq t < T_{m+1}\}} i_m + I_{\{t \geq T_\infty\}} i_\infty \quad \text{for } t \geq 0,$$

where I_D denotes the indicator function of a set D. The process after T_∞ is regarded to be absorbed in the state i_∞. Hence, we write $q(i_\infty|i_\infty, a_\infty) = 0, c(i_\infty, a_\infty) = 0, A(i_\infty) := \{a_\infty\}, A_\infty := A \cup \{a_\infty\}$, where a_∞ is an isolated point. Let $\mathcal{F}_t := \sigma(\{T_m \leq s, X_m = i\} : i \in S, s \leq t, m \geq 0)$ for $t \geq 0$, $\mathcal{F}_{s-} := \bigvee_{0\leq t<s} \mathcal{F}_t$, and $\mathcal{P} := \sigma(\{D \times \{0\}, D \in \mathcal{F}_0\} \cup \{D \times (s, \infty), D \in \mathcal{F}_{s-}, s > 0\})$ which denotes the σ-algebra of predictable sets on $\Omega \times [0,\infty)$ related to $\{\mathcal{F}_t\}_{t\geq 0}$.

Now we introduce the definition of a randomized Markov policy below.

Definition 2.1. A $\mathcal{P}$-measurable transition probability $\pi(\cdot|\omega, t)$ on $(A_\infty, \mathcal{B}(A_\infty))$, concentrated on $A(\xi_{t-}(\omega))$ is called a randomized Markov policy if there exists a kernel φ on A_∞ given $S_\infty \times [0,\infty)$ such that $\pi(\cdot|\omega, t) = \varphi(\cdot|\xi_{t-}(\omega), t)$. A policy π is said to be deterministic stationary if there exists a function f on S_∞ satisfying $f(i) \in A(i)$ for all $i \in S_\infty$ and $\pi(\cdot|\omega, t) = \delta_{f(\xi_{t-}(\omega))}(\cdot)$, where $\delta_x(\cdot)$ is the Dirac measure concentrated at the point x.

The set of all randomized Markov policies and the set of all deterministic stationary policies are denoted by Π and F, respectively.

For any initial state $i \in S$ and any $\pi \in \Pi$, Theorem 4.27 in [8] gives the existence of a unique probability measure P_i^π on $(\Omega, \mathcal{F})$. Moreover, the expectation operator with respect to P_i^π is denoted by E_i^π.

Fix an arbitrary risk-sensitivity coefficient $\lambda > 0$ throughout this paper. For any $i \in S$ and $\pi \in \Pi$, the risk-sensitive average cost criterion is defined by

$$J(i,\pi) = \limsup_{T\to\infty} \frac{1}{\lambda T} \ln E_i^\pi \left[e^{\lambda \int_0^T \int_A c(\xi_t, a)\pi(da|\xi_t, t)dt} \right].$$

The corresponding optimal value function is given by

$$J^*(i) := \inf_{\pi\in\Pi} J(i,\pi) \quad \text{for all } i \in S.$$

Definition 2.2. A policy $\pi^* \in \Pi$ is said to be optimal if $J(i,\pi^*) = J^*(i)$ for all $i \in S$.

The main goals of this paper are to give the conditions for the existence of optimal policies and to develop a new approach to establish the existence of a solution to the risk-sensitive average cost optimality equation.

3. The optimality conditions and main results

In this section, we establish the existence of a solution to the risk-sensitive average cost optimality equation, from which the existence of optimal policies can be shown. To this end, we first introduce the following optimality conditions.

Assumption 3.1. (i) For each $i \in S$, the set $A(i)$ is compact.

(ii) For each $i, j \in S$, the functions $c(i,a)$ and $q(j|i,a)$ are continuous in $a \in A(i)$.

(iii) For each $f \in F$, the corresponding continuous-time Markov chain $\{\xi_t, t \geq 0\}$ is irreducible, which means that for any two states $i \neq j$, there exist different states $j_1 = i, j_2, \ldots, j_m$ such that $q(j_2|j_1, f) \cdots q(j|j_m, f) > 0$, where $q(j|i,f) := q(j|i, f(i))$.

Remark 3.1. Assumption 3.1(i) and (ii) are the standard continuity and compactness conditions which have been widely used in CTMDPs; see, for instance, [5,6,10] and the references therein. Moreover, Assumption 3.1(i) and the Tychonoff theorem imply that F is compact and metrizable. Assumption 3.1(iii) is the so-called irreducibility condition which is commonly used in the average cost criterion; see, for instance, [6] for the expected average case and [5] for the risk-sensitive average case.

In order to prove the existence of optimal policies, we introduce the following notation.

For any fixed state $z \in S$, set $\tau_z := \inf\{t \geq T_1 : \xi_t = z\}$ with $\inf \emptyset := \infty$. For each $i \in S$ and $f \in F$, let $c(i,f) := c(i, f(i))$. Below we introduce a risk-sensitive first passage optimization problem which has not been discussed in the existing literature. For each $g \in \mathbb{R} := (-\infty, \infty)$, $i \in S$ and $f \in F$, we define

$$h_g(i,f) := \frac{1}{\lambda} \ln E_i^f \left[e^{\lambda \int_0^{\tau_z} (c(\xi_t, f) - g)dt} \right] \quad \text{and}$$
$$h_g^*(i) := \inf_{f\in F} h_g(i,f). \tag{3.1}$$