CrossMark

J. Appl. Prob. **58,** 523–550 (2021)
*doi:*10.1017/jpr.2020.105

RISK-SENSITIVE AVERAGE CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH UNBOUNDED TRANSITION AND COST RATES

XIN GUO,* *Tsinghua University*
YONGHUI HUANG,** *Sun Yat-Sen University*

Abstract

This paper considers risk-sensitive average optimization for denumerable continuous-time Markov decision processes (CTMDPs), in which the transition and cost rates are allowed to be unbounded, and the policies can be randomized history dependent. We first derive the multiplicative dynamic programming principle and some new facts for risk-sensitive finite-horizon CTMDPs. Then, we establish the existence and uniqueness of a solution to the risk-sensitive average optimality equation (RS-AOE) through the results for risk-sensitive finite-horizon CTMDPs developed here, and also prove the existence of an optimal stationary policy via the RS-AOE. Furthermore, for the case of finite actions available at each state, we construct a sequence of models of finite-state CTMDPs with optimal stationary policies which can be obtained by a policy iteration algorithm in a finite number of iterations, and prove that an average optimal policy for the case of infinitely countable states can be approximated by those of the finite-state models. Finally, we illustrate the conditions and the iteration algorithm with an example.

Keywords: Continuous-time Markov decision process; unbounded transition and cost rate; risk-sensitive average optimality equation; optimal policy; finite approximation

2010 Mathematics Subject Classification: Primary 90C40
Secondary 60J27

1. Introduction

The study of risk-sensitive Markov decision processes in discrete time started in the 1970s [16], and there have since been a lot of developments; see, e.g., [2, 3, 4, 5, 19, 20]. In comparison, there has been limited work on risk-sensitive continuous-time Markov decision processes (CTMDPs). For background on risk-sensitive CTMDPs, see [7, 11, 22] for the risk-sensitive (infinite-horizon) discounted criterion, [7, 22, 23, 25, 26, 27] for the risk-sensitive (long-run) average criterion, [7, 9, 14, 17, 18, 23, 24] for the risk-sensitive finite-horizon criterion, and [8, 28] for the risk-sensitive total cost criterion. In this paper we will further consider the risk-sensitive average criterion for CTMDPs studied in [7, 22, 23, 25, 26, 27]. More precisely, [7] proved the existence of an optimal policy under the geometric ergodicity assumption and Doeblin-type conditions for the small cost function and bounded transition rates, and provided a policy improvement algorithm for the case of finite states and actions, but they did

Received 15 January 2020; revision received 16 November 2020.
* Postal address: School of Economics and Management, Tsinghua University, Beijing, China.
** Postal address: School of Mathematics, Sun Yat-Sen University, and Guangdong Province Key Laboratory of Computational Science, Sun Yat-Sen University, Guangzhou, China. Email address: hyongh5@mail.sysu.edu.cn

© The Author(s), 2021. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

not establish a characterization of the optimality equation. Using the multiplicative dynamic programming principle, [22] proved the existence and a characterization of an optimal policy under the positive recurrent assumption for a near-monotone cost function and bounded transition rates, and [23] under the geometric ergodicity conditions of the state dynamics for small running cost and bounded transition rates. Wei and Chen [25, 26] considered the case of finite states under the ergodicity condition. The arguments for the existence of an average optimal deterministic stationary policy in [7, 22, 23, 25, 26] are based on the ergodicity conditions and assumptions of bounded transition rates. For a general case of unbounded cost and transition rates, [27] proved the existence of an optimal policy under irreducibility and Doeblin's conditions, using the risk-sensitive average optimality inequality. Moreover, only stationary policies were considered in [25], whereas the general class of Markov policies was taken into account in [7, 22, 23, 25, 26, 27].

In the present work we will further study the general case that the transition and cost rates are allowed to be unbounded. We not only establish the existence and uniqueness of a solution to the risk-sensitive average optimality equation (RS-AOE) and prove the existence of an optimal stationary policy via the RS-AOE, but also provide an approximation algorithm for computing average optimal policies. Our arguments are inspired by those of [23] in constructing a solution to the RS-AOE, which is more specific than the risk-sensitive average optimality inequality in [27]. Apart from allowing unbounded transition and cost rates, we consider a more general class of randomized history-dependent policies, and do not require the condition that the controlled process is irreducible under each deterministic stationary policy, where the irreducibility condition was also assumed in [7, 22, 23, 25, 26, 27]. Some of our conditions are extensions of those in [7, 22, 23] and different from those in [27]; see Remarks 3.1, 3.3, 7.2, and 7.3 for more details.

First, under the milder conditions from [9, 12, 14], as in the arguments of [7, 23] we derive the multiplicative dynamic programming principle and obtain some interesting results, such as the existence and uniqueness of a solution to the optimality equation and properties of the value function for risk-sensitive finite-horizon CTMDPs with unbounded transition and cost rates (Theorem 3.1 and Lemma 3.2). Second, using the multiplicative dynamic programming principle and properties developed here for risk-sensitive finite-horizon CTMDPs, by the method in [7, 23] we obtain the existence of a solution to the RS-AOE and its characterization for risk-sensitive average CTMDPs (Theorems 4.1 and 6.2). Third, to prove the existence of a risk-sensitive average optimal policy via the RS-AOE, we provide some new and verifiable sufficient conditions under which we prove that the solution to the RS-AOE is bounded away from zero (Proposition 5.1), and under which we establish the existence of a risk-sensitive average optimal stationary policy over the class of all randomized history-dependent policies (Theorem 5.1). Fourth, after showing the solvability of the multiplicative Poisson equation, we construct a sequence of models of finite-state CTMDPs for which the corresponding optimal stationary policies can be obtained by the policy iteration algorithm in a finite number of iterations (Theorem 6.1), and further prove that an average optimal policy for the case of infinitely countable states can be approximated by those of the finite-state models (Theorem 6.2). Finally, in order to illustrate our conditions and the main results, we present an example on a controlled population with *unbounded* transition rates and cost functions [1], in which all of our conditions are satisfied, while some of the conditions in [7, 22, 23, 25, 26, 27] fail to hold; see Remarks 7.2 and 7.3. Also, we conduct a numerical experiment to show the effectiveness of the finite-approximation policy iteration algorithm.

The rest of the paper is organized as follows. We describe the optimization problem in Section 2, and derive the multiplicative dynamic programming principle and some new facts

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

for risk-sensitive finite-horizon CTMDPs in Section 3, where some comments on our conditions are given by Remarks 3.1, 3.3, and 3.4 and Example 3.1. We establish the existence of an optimal policy in Section 5, after establishing the RS-AOE in Section 4. We provide a finite-approximation algorithm for computing optimal policies and prove its convergence in Section 6, and illustrate our main results with an example in Section 7. Finally, we provide an appendix for the proofs of some preliminaries in Section 8.

2. The risk-sensitive average optimality problem

The model of CTMDPs is a four-tuple, $\mathbb{M} := \{S, A(i), c(i, a), q(j \mid i, a)\}$, consisting of the following elements:

- a denumerable set S, called the state space, whose elements are referred to as states of a system (for simplicity, without loss of generalization, let $S := \{0, 1, \ldots, i, \ldots\}$);
- the action sets $A(i)$, each of which denotes the set of actions available to a controller when the system is in state $i \in S$, and is assumed to be a compact subset of a Borel space A;
- the running cost rate function $c(i, a)$ that is assumed to be real valued, Borel measurable, and bounded from below on K, where $K := \{(i, a) \mid i \in S, a \in A(i)\}$;
- the transition rates $q(j \mid i, a)$, a Borel-measurable signed kernel on S given K, taking nonnegative values for all $j, i \in S$ with $j \neq i$, being conservative in the sense of $q(S \mid i, a) \equiv 0$, and stable in that of

$$q^*(i) := \sup_{a \in A(i)} q(i, a) < \infty \qquad \text{for all } i \in S, \tag{1}$$

where $q(i, a) := -q(i \mid i, a) \geq 0$ for all $(i, a) \in K$.

To formalize this, below we show the construction of CTMDPs under possibly randomized history-dependent policies. To do so, we introduce some notation. Let $S_\Delta := S \cup \{\Delta\}$ (with some $\Delta \notin S$), $\Omega^0 := (S \times (0, \infty))^\infty$,

$$\Omega := \Omega^0 \cup \{(i_0, \theta_1, i_1, \ldots, \theta_k, i_k, \infty, \Delta, \infty, \ldots) \mid i_0, i_l \in S, \theta_l \in (0, \infty) \text{ for each } 1 \leq l \leq k, k \geq 1\},$$

and let $\mathcal{F}$ be the Borel σ-algebra on Ω. Then we obtain the measurable space $(\Omega, \mathcal{F})$. For each $k \geq 0$ and $\omega := (i_0, \theta_1, i_1, \ldots, \theta_k, i_k, \ldots) \in \Omega$, let $h_k(\omega) := (i_0, \theta_1, i_1, \ldots, \theta_k, i_k)$ denote the k-component internal history, and define $T_0(\omega) := 0$, $T_{k+1}(\omega) := \theta_1 + \theta_2 + \cdots + \theta_{k+1}$, $X_k(\omega) := i_k$. In what follows, the argument ω is always omitted. Let $T_\infty := \lim_{k \to \infty} T_k$, and define the continuous-time process $\{x_t\}$ by

$$x_t := \sum_{k \geq 0} \mathbf{1}_{\{T_k \leq t < T_{k+1}\}} X_k + \Delta \mathbf{1}_{\{t \geq T_\infty\}} \qquad \text{for } t \geq 0. \tag{2}$$

Here and below, $\mathbf{1}_E$ stands for the indicator function on any set E, and we use the notation that $0 + z = z$ and $0z = 0$ for all $z \in S_\Delta$.

From (2), we see that T_k ($k \geq 1$) denotes the kth jump moment of $\{x_t\}$, X_{k-1} is the state of the process on $[T_{k-1}, T_k)$, and $\Theta_k := T_k - T_{k-1}$ plays the role of sojourn time at state X_{k-1}. We do not intend to consider the controlled process after moment T_∞, and thus view it to

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

be absorbed in the state Δ. Hence, we write $q(\cdot \mid \Delta, a_\Delta) \equiv 0$, $c(\Delta, a_\Delta) \equiv 0$, $A(\Delta) := \{a_\Delta\}$, $A_\Delta := A \cup \{a_\Delta\}$, where a_Δ is an isolated point.

To state the optimality problem, we need to introduce the concept of a policy [9, 15, 21].

Definition 2.1. A *randomized history-dependent policy* π is determined by a sequence $\{\pi^k, k \geq 0\}$ such that

$$\pi(\mathrm{d}a \mid \omega, t) = \mathbf{1}_{\{0\}}(t)\pi_0(\mathrm{d}a \mid i_0, 0) + \sum_{k\geq 0} \mathbf{1}_{\{T_k < t \leq T_{k+1}\}} \pi^k(\mathrm{d}a \mid i_0, \theta_1, i_1, \ldots, \theta_k, i_k, t - T_k) + \mathbf{1}_{\{t \geq T_\infty\}} \delta_{a_\Delta}(\mathrm{d}a),$$

where $\pi_0(\mathrm{d}a \mid i_0, 0)$ is a stochastic kernel on A given S such that $\pi_0(A(i_0) \mid i_0, 0) = 1$, π^k $(k \geq 0)$ are stochastic kernels on A given $(S \times (0, \infty))^{k+1}$ such that $\pi^k(A(i_k) \mid i_0, \theta_1, i_1, \ldots, \theta_k, i_k, t - T_k) = 1$, and $\delta_{a_\Delta}(\mathrm{d}a)$ denotes the Dirac measure at the point a_Δ.

A policy π is called randomized Markovian if each π^k has the form $\pi^k(\mathrm{d}a \mid i_k, t)$ (independent of the histories $(i_0, \theta_1, i_1, \ldots, \theta_k)$), and it will be denoted by $\pi_t(\mathrm{d}a \mid \cdot)$. A randomized Markov policy $\pi_t(\mathrm{d}a \mid \cdot)$ is called deterministic Markovian whenever there exists an A-valued Borel-measurable function $f_t(i)$ on $[0, \infty) \times S$ such that $\pi_t(\mathrm{d}a \mid i)$ is a Dirac measure concentrated at $f_t(i)$ for each $(t, i) \in [0, \infty) \times S$. Such a Markov policy will be denoted by f_t for simplicity. A Markov policy f_t is called a (deterministic) stationary one if f_t is independent of time t. Such a stationary policy will be denoted by f for simplicity.

We denote by Π the set of all randomized history-dependent policies, by $\Pi_{\mathrm{m}}^{\mathrm{r}}$ the set of all randomized Markov policies, by $\Pi_{\mathrm{m}}^{\mathrm{d}}$ the set of all (deterministic) Markov policies, and by F the set of all stationary policies. Obviously, $\Pi \supset \Pi_{\mathrm{m}}^{\mathrm{r}} \supset \Pi_{\mathrm{m}}^{\mathrm{d}} \supset F$.

For any initial distribution γ on S and policy $\pi \in \Pi$, we introduce the unique probability measure $\mathbb{P}_\gamma^\pi$ on $(\Omega, \mathcal{F})$ as in [12, 13, 28]. Let $\mathbb{E}_\gamma^\pi$ be its corresponding expectation operator. In particular, $\mathbb{E}_\gamma^\pi$ and $\mathbb{P}_\gamma^\pi$ will be respectively written as $\mathbb{E}_i^\pi$ and $\mathbb{P}_i^\pi$ when γ is the Dirac measure located at a state i in S.

As in [7, 22, 23, 25], the risk-sensitive average criterion of CTMDPs is defined by

$$J(i, \pi) := \limsup_{T\to\infty} \frac{1}{T} \ln \mathbb{E}_i^\pi \left[\exp\left\{ \int_0^T \int_A c(x_t, a)\pi(\mathrm{d}a \mid \omega, t)\, \mathrm{d}t \right\} \right]$$

for each $i \in S$ and $\pi \in \Pi$.

Definition 2.2. A policy $\pi^* \in \Pi$ is said to be risk-sensitive (R-S) average optimal if $J(i, \pi^*) = \inf_{\pi\in\Pi} J(i, \pi)$ for all $i \in S$.

This paper will give new conditions for the existence and computation of R-S average optimal stationary policies.

Since $c(i, a)$ is bounded below (i.e. $c(i, a) \geq L$ for all $(i, a) \in K$ and some constant L), we have $c(i, a) + |L| \geq 0$ on K, and

$$\limsup_{T\to\infty} \frac{1}{T} \ln \mathbb{E}_i^\pi \left[\exp\left\{ \int_0^T \int_A [c(x_t, a) + |L|]\pi(\mathrm{d}a \mid \omega, t)\, \mathrm{d}t \right\} \right] = J(i, \pi) + |L|.$$

Thus, adding a constant to all the costs $c(i, a)$ will affect all policies identically under the R-S average criterion. Therefore, without loss of generality, we will further assume that $c(i, a)$ is *nonnegative* on K.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

3. Preliminaries

To prove the existence of risk-sensitive average optimal policies, we need to develop some preliminary facts concerning risk-sensitive finite-horizon CTMDPs, some of which are from [9] and some of which are new.

3.1. Non-explosion and finiteness

Since the transition and cost rates may be unbounded, we need the following conditions from [9, 12, 15, 23] to guarantee the non-explosion of $\{x_t, t \geq 0\}$ and the finiteness of $J(i, \pi)$.

Assumption 3.1. *There exist real-valued functions $V_0 \geq 1$ and $\delta \geq 0$ on S, positive constants b_0 and M_0, and a state $i_0 \in S$ such that:*

(i) $\sum_{j \in S} V_0(j) q(j \mid i, a) \leq -\delta(i) V_0(i) + b_0 \mathbf{1}_{\{i_0\}}(i)$ *for all* $(i, a) \in K$;

(ii) $q^*(i) \leq M_0 V_0(i)$ *for all* $i \in S$, *where* $q^*(i)$ *is as in (1);*

(iii) $\delta_* := \inf_{i \neq i_0} \delta(i) > 0$ *and* $0 \leq c(i, a) \leq \delta(i) \leq \sqrt{\ln V_0(i)}$ *for all* $(i, a) \in K$.

Remark 3.1.

(a) Although the indicator function $\mathbf{1}_{\{i_0\}}$ in Assumption 3.1(i) is a modification of the indicator function $\mathbf{1}_C$ in [23] with a finite subset C of S, such a modification is required; see Remark 3.4 for more details. Assumption 3.1(i) is an extension of Assumption (A2) in [23], Assumption (A5) in [7], Assumption 3.1(i) in [27], and Assumption 7.1 in [10], from a constant δ in [7, 10, 23, 27] to a function $\delta(i)$ on S here. Thus, it is satisfied for the examples in [7, 10] and will be verified with other examples below.

(b) Assumption 3.1(iii) is new and ensures the finiteness of

$$\mathbb{E}_i^\pi \left[\exp \left\{ \int_0^{\tau_{i_0}} \int_A c(x_v, a) \pi(\mathrm{d}a \mid \omega, v) \, \mathrm{d}v \right\} \right],$$

where $\tau_{i_0} := \min\{t \geq 0 \mid x_t = i_0\}$ denotes the first passage time to i_0. Since the function $c(i, a)$ here is allowed to be unbounded, Assumption 3.1(iii) is an extension of the so-called smallness condition (i.e. $\sup_{(i,a) \in K} |c(i, a)| < \delta_*$) in [7, 23] and different from the near-monotone condition in [22]; see Example 7.1 and Remark 7.3 for details.

(c) Although Assumption 3.1(i) is slightly stronger than Assumption 3.1(i) in [27], Assumptions 3.3(ii) and 3.4(ii) in [27] are not required here.

To characterize risk-sensitive CTMDPs over the class of all history-dependent policies, we need an extension of the Feymanm–Kac formula in [9] from the Markovian case to a more general case, which is based on the following conditions.

Assumption 3.2. *There exist a real-valued function $V_1 \geq 1$ on S and positive constants ρ_1, b_1, and M_1 such that:*

(i) $\sum_{j \in S} V_1^2(j) q(j \mid i, a) \leq \rho_1 V_1^2(i) + b_1$ *for all* $(i, a) \in K$;

(ii) $V_0^2(i) \leq M_1 V_1(i)$ *for all* $i \in S$, *where* V_0 *comes from Assumption 3.1.*

It can be proved that Assumption 3.2 is not required when the transition rates are bounded (i.e. $\sup_{j \in S, (i,a) \in K} |q(j \mid i, a)| < \infty$).

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

Assumption 3.3.

(i) For any fixed $i, j \in S$, $q(j \mid i, a)$ and $c(i, a)$ are continuous in $a \in A(i)$;

(ii) For any given $i \in S$, the convergence of $\sum_{j \in S} V_0(j) q(j \mid i, a)$ holds uniformly in $a \in A(i)$.

Remark 3.2. Assumption 3.3(ii) is an extension of (A3) in [7]. Moreover, under Assumption 3.3(i), Assumption 3.3(ii) is fulfilled automatically when $\sum_{j \in S(i)} q(j \mid i, a) = 0$ for $a \in A(i)$, where $S(i)$ is a finite subset of S and can depend on $i \in S$. Assumption 3.3 implies that $\sum_{j \in S} V_0(j) q(j \mid i, a)$ is continuous in $a \in A(i)$, and so is slightly stronger than the continuity–compactness conditions [7, 22, 23]. In fact, it is required not only for the facts in [9] but also for the existence of $\varphi'(t, i)$ at all $t \geq 0$ in Theorem 3.1 (instead of at almost everywhere $t \geq 0$ in [23]); see Remark A.1 for more details.

Lemma 3.1. *If Assumptions 3.1 and 3.2 are satisfied, then the following assertions hold:*

(a) For $i \in S$ and $\pi \in \Pi$,

$$\mathbb{E}_i^\pi \left[\exp \left\{ \int_0^{\tau_{i_0}} \int_A c(x_v, a) \pi(\mathrm{d}a \mid \omega, v) \, \mathrm{d}v \right\} \right] \leq \mathbb{E}_i^\pi \left[\exp \left\{ \int_0^{\tau_{i_0}} \delta(x_v) \, \mathrm{d}v \right\} \right] \leq V_0(i).$$

(b) $\mathbb{E}_i^\pi [\tau_{i_0}] \leq \frac{V_0(i)}{\delta_*}$ and $\mathbb{P}_i^\pi (\tau_{i_0} < \infty) = 1$, for all $i \in S$ and $\pi \in \Pi$.

Proof. See Appendix A. □

Remark 3.3. In this paper, i_0 in Lemma 3.1 comes directly from Assumption 3.1(i). However, in [23], the corresponding reference state i_0 is determined by the condition $V(i_0) \geq 1 + \frac{b}{\delta}$, where the function V and constants δ as well as b are from [23, Assumption A2], corresponding to the V_0 and $\frac{\delta(i)}{2}$ as well as b_0 in Assumption 3.1 here. Example 3.1 shows that the condition $V(i_0) \geq 1 + \frac{b}{\delta}$ actually cannot be used to get a reference state.

Example 3.1. Let $S := \{0, 1, 2\}$, all $A(i)$ be singleton sets; then there is a unique stationary policy, say f. Moreover, let $q(0 \mid 0, f(0)) = -1$, $q(1 \mid 0, f(0)) = 1$, $q(2 \mid 0, f(0)) = 0$; $q(0 \mid 1, f(1)) = 8$, $q(1 \mid 1, f(1)) = -9$, $q(2 \mid 1, f(1)) = 1$; $q(0 \mid 2, f(2)) = 0$, $q(1 \mid 2, f(2)) = 8$, $q(2 \mid 2, f(2)) = -8$.

Obviously, Assumption A2 in [23] is satisfied for $V(i) := 3 + i$ for $i \in S$, $C := \{0, 1\}$ (or $\{0\}$), $\delta := \frac{1}{2}$, $b := 4$, which implies that $1 + \frac{b}{\delta} = 9 > V(i)$ for all $i \in S$. Thus, if we follow the statement in [23], this example does not have any reference state.

However, since Assumptions 3.1 and 3.2 are also satisfied for $V_0(i) = V_1(i) := 3 + i$ for $i \in S$ and $i_0 = 0$, the state 0 is actually a reference state for this paper.

3.2. Finite-horizon optimality

In order to deal with risk-sensitive average optimality, we need some notation and facts on risk-sensitive finite-horizon optimality.

For any $\pi \in \Pi$, $t \geq 0$, $i \in S$, the t-horizon risk-sensitive criterion

$$\varphi(t, i, \pi) := \mathbb{E}_i^\pi \left[\exp \left\{ \int_0^t \int_A c(x_s, a) \pi(\mathrm{d}a \mid \omega, s) \, \mathrm{d}s \right\} \right]$$

is well defined. Then, let $\varphi(t, i) := \inf_{\pi \in \Pi} \varphi(t, i, \pi)$ for $i \in S$, which is called the value function of the t-horizon risk-sensitive criterion. Since c is nonnegative, $\varphi(t, i)$ is increasing in $t \geq 0$, $\varphi(0, i) = 1$, and $\varphi(t, i) \geq 1$ for all $t \geq 0$.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

To state the results of the finite-horizon case, let us introduce some additional notations. Given any function $W \geq 1$ on S, a function $u(t, i)$ on $[0, T] \times S$ is called W-bounded if the W-weighted norm of $u(t, i)$, $\|u\|_W := \sup_{(t,i)\in[0,T]\times S} \frac{|u(t,i)|}{W(i)}$, is finite. Let $\mathbb{B}_W([0, T] \times S)$ be the Banach space of all W-bounded measurable functions $u(t, i)$ on $[0, T] \times S$. For each measurable function $u(t, i)$ on $[0, T] \times S$, if $u(\cdot, i)$ is absolutely continuous in $t \in [0, T]$, then we put a measurable function $u'(t, i)$ on $[0, T] \times S$ such that $u(s, i) - u(0, i) = \int_0^s u'(t, i)\, \mathrm{d}t$ for each $i \in S$ and $s \in [0, T]$. This also means that $u'(t, i)$ is the derivative with respect to time t in $[0,T]$. For V_0 and V_1 from Assumptions 3.1 and 3.2, we denote by $\mathbb{B}_{V_0,V_1}([0, T] \times S)$ the collection of functions $u(t, i)$ in $\mathbb{B}_{V_0}([0, T] \times S)$ such that each $u(t, i)$ (for any given $i \in S$) is absolutely continuous in $t \in [0, T]$ and admits a V_1-bounded $u'(t, i)$.

Theorem 3.1. *(Finite-horizon optimality.) Under Assumptions 3.1, 3.2, and 3.3, for any $T > 0$, the following assertions hold.*

(a) The value function $\varphi(t, i)$ is the unique solution in $\mathbb{B}_{V_0,V_1}([0, T] \times S)$ of the following risk-sensitive finite-horizon optimality equation:

$$\begin{cases} \varphi'(t, i) = \inf_{a\in A(i)} [c(i, a)\varphi(t, i) + \sum_{j\in S} \varphi(t, j)q(j \mid i, a)], \\ \varphi(0, i) = 1, \end{cases}$$

for each $(t, i) \in [0, T] \times S$, and $\varphi'(t, i)$ is continuous on $[0, T] \times S$.

(b) There exists a Markov policy $\pi^* = f_t^* \in \Pi_{\mathrm{m}}^{\mathrm{d}}$ such that $\varphi'(t, i) = c(i, f_t^*(i))\varphi(t, i) + \sum_{j\in S} \varphi(t, j)q(j \mid i, f_t^*(i))$ for all $(t, i) \in [0, T] \times S$, and $\varphi(t, i) = \varphi(t, i, \pi^*) = \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \varphi(t, i, \pi)$ for all $i \in S$ and $t \in [0, T]$.

(c) (The multiplicative dynamic programming principle.) For any subset D of S, let $\tau_D := \inf\{s \geq 0 \mid x_s \in D\}$ be the hitting time of the process $\{x_t\}$ to a set D. Then, for all $i \in S$ and $t \geq 0$, the value function $\varphi(t, i)$ can be represented as

$$\varphi(t, i) = \inf_{f_t\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^{\pi} \left[\exp\left\{ \int_0^{t\wedge\tau_D} c(x_v, f_v(x_v))\, \mathrm{d}v \right\} \varphi(t - t \wedge \tau_D, x_{t\wedge\tau_D}) \right].$$

Proof. See Appendix A. □

Lemma 3.2. *Under Assumptions 3.1, 3.2, and 3.3, the following assertions hold:*

(a) $\frac{\varphi(t,i)}{\varphi(t,i_0)} \leq V_0(i)$ for all $i \in S$ and $t \geq 0$;

(b) $\varphi'(t, i) \geq 0$ and $\sup_{t\geq 0} \frac{\varphi'(t,i)}{\varphi(t,i_0)} \leq L(i)$ for $i \in S$ and $t \geq 0$, where $L(i) := 2V_0(i)q^*(i) + b_0$.

Proof. See Appendix A. □

Remark 3.4. If the function $\mathbf{1}_{\{i_0\}}$ in Assumption 3.1(i) is replaced with the indicator function $\mathbf{1}_C$ of a finite subset C of S, as in the proof of Lemma 3.2 in the appendix, we can prove that, for all $t \geq 0$,

$$\begin{aligned} &\mathbb{E}_i^{\pi} \left[\exp\left\{ \int_0^{\tau_C} \delta(x_v)\, \mathrm{d}v \right\} \right] \leq V_0(i), \\ &\varphi(t, i) \leq \mathbb{E}_i^{\pi} \left[\exp\left\{ \int_0^{\tau_C} c(x_v, \pi_v)\, \mathrm{d}v \right\} \right] \varphi(t, i^*(t)), \end{aligned} \tag{3}$$

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

where the states $i^*(t)$, depending on $t \geq 0$, are determined by $\varphi(t, i^*(t)) := \max_{i \in C} \varphi(t, i)$. However, we cannot prove Lemma 3.2 by (3). In fact, since $\tau_{i_0} \geq \tau_C$ (by $i_0 \in C$) and the $i^*(t)$ may change with t, from (3) we cannot establish the existence of some fixed $i_0 \in C$ such that, for all $t \geq 0$,

$$\mathbb{E}_i^\pi \left[\exp\left\{ \int_0^{\tau_{i_0}} \delta(x_v)\, \mathrm{d}v \right\} \right] \leq V_0(i), \qquad \frac{\varphi(t, i)}{\varphi(t, i_0)} \leq \mathbb{E}_i^\pi \left[\exp\left\{ \int_0^{\tau_C} c(x_v, \pi_v)\, \mathrm{d}v \right\} \right],$$

which are required in [23].

4. The risk-sensitive average optimality equation

In this section we establish the existence of a solution to the RS-AOE for CTMDPs. To begin with, we need some notation. For each $n \geq 1$, let

$$c_n(i, a) := \begin{cases} c(i, a), & \text{for } i \in \{0, \ldots, n\},\ a \in A(i), \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Using Theorem 3.1 and Lemma 3.2, and the method in [23], we extend the corresponding results in [23] to the case of unbounded transition and cost rates.

Theorem 4.1. *Under Assumptions 3.1, 3.2, and 3.3, for each $n \geq i_0$:*

(a) There exists a solution (ρ_n, ψ_n) in $[0, L(i_0)] \times \mathbb{B}_{V_0}(S)$ to

$$\rho_n \psi_n(i) = \inf_{a \in A(i)} \left\{ c_n(i, a)\psi_n(i) + \sum_{j \in S} \psi_n(j) q(j \mid i, a) \right\}, \quad \psi_n(i_0) = 1, \quad i \in S. \tag{5}$$

(b) $\rho_n \leq \inf_{\pi \in \Pi} J(i, \pi)$ for all $i \in S$.

(c) There is a policy f_n^* with $f_n^* \in F$ such that, for all $i \in S$,

$$\begin{aligned} \psi_n(i) &= \mathbb{E}_i^{f_n^*} \left[\exp\left\{ \int_0^{\tau_{i_0}} (c_n(x_t, f_n^*(x_t)) - \rho_n)\, \mathrm{d}t \right\} \right] \\ &= \inf_{\pi \in \Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^\pi \left[\exp\left\{ \int_0^{\tau_{i_0}} (c_n(x_t, \pi_t) - \rho_n)\, \mathrm{d}t \right\} \right]. \end{aligned}$$

Proof. See Appendix A. □

By Theorem 4.1, we can establish the RS-AOE in the following theorem.

Theorem 4.2. *Under Assumptions 3.1, 3.2, and 3.3, the following hold.*

(a) There exists a solution (ρ^*, ψ^*) in $[0, L(i_0)] \times \mathbb{B}_{V_0}(S)$ to the RS-AOE

$$\psi^*(i_0) = 1, \quad \rho^* \psi^*(i) = \inf_{a \in A(i)} \left\{ c(i, a)\psi^*(i) + \sum_{j \in S} \psi^*(j) q(j \mid i, a) \right\} \quad \text{for all } i \in S.$$

(b) $\rho^* \leq \inf_{\pi \in \Pi} J(i, \pi)$ for all $i \in S$.

(c) There exists some policy f^* with $f^* \in F$ such that:

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

(c_1) $\rho^*\psi^*(i) = c(i, f^*(i))\psi^*(i) + \sum_{j\in S} \psi^*(j)q(j \mid i, f^*(i))$ *for all* $i \in S$;
(c_2) *for all* $i \in S$,

$$\psi^*(i) = \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, \pi_t) - \rho^*)\,\mathrm{d}t\right\}\right]$$
$$= \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\right].$$

Proof. (a) and (b) Take (ρ_n, ψ_n) as in Theorem 4.1 Then, since $\rho_n \in [0, L(i_0)]$ for all $n \geq i_0$, there exist a subsequence $\{n_1\}$ of $\{n, n \geq i_0\}$ and a constant $\rho^* \in [0, L(i_0)]$ such that $\rho^* = \lim_{n_1\to\infty} \rho_{n_1}$, and $\rho^* \leq \inf_{\pi\in\Pi} J(i, \pi)$ (by Theorem 4.1(b)). Furthermore, since $\psi_n \in \mathbb{B}_{V_0}(S)$, $\psi_n(i_0) = 1$ for all $n \geq i_0$, and S is denumerable, the diagonalization argument ensures the existence of a subsequence $\{n_2\}$ of $\{n_1\}$ and a function $\psi^* \in \mathbb{B}_{V_0}(S)$ satisfying $\rho^* = \lim_{n_2\to\infty} \rho_{n_2}$, $\psi^*(i) = \lim_{n_2\to\infty} \psi_{n_2}(i)$ for all $i \in S$, $\psi^*(i_0) = 1$. Then, replacing n in (5) with n_2 and letting $n_2 \to \infty$, we get

$$\rho^*\psi^*(i) = \inf_{a\in A(i)}\left\{c(i, a)\psi^*(i) + \sum_{j\in S} \psi^*(j)q(j \mid i, a)\right\} \qquad \text{for all } i \in S.$$

Thus, we complete the proof of (a) and (b).

(c) For any given $\pi \in \Pi_{\mathrm{m}}^{\mathrm{d}}$, since $c_{n_2} \leq c$ and $\rho_{n_2} \geq 0$, Theorem 4.1(c) gives, for all $n_2 \geq 1$,

$$\psi_{n_2}(i) \leq \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} (c_{n_2}(x_t, \pi_t) - \rho_{n_2})\,\mathrm{d}t\right\}\right]$$
$$\leq \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} c(x_t, \pi_t)\,\mathrm{d}t\right\}\right] \leq V_0(i).$$

Thus, using the dominated convergence theorem and letting $n_2 \to \infty$, we have $\psi^*(i) \leq \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, \pi_t) - \rho^*)\,\mathrm{d}t\right\}\right]$, and hence

$$\psi^*(i) \leq \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, \pi_t) - \rho^*)\,\mathrm{d}t\right\}\right] \qquad \text{for all } i \in S. \tag{6}$$

Taking $f_{n_2}^*(i)$ as the minimum of (5),

$$\rho_n\psi_{n_2}(i) = c(i, f_{n_2}^*(i))\psi_{n_2}(i) + \sum_{j\in S} \psi_{n_2}(j)q(j \mid i, f_{n_2}^*(i)) \qquad \text{for all } i \in S. \tag{7}$$

Since $f_{n_2}^*(i) \in A(i)$ belongs to the compact $A(i)$ (for each $i \in S$) and S is denumerable, there exists a subsequence $\{n_3\}$ of $\{n_2\}$ and $f^* \in F \subset \Pi_{\mathrm{m}}^{\mathrm{d}}$ such that $c(i, f^*(i)) = \lim_{n_3\to\infty} c(i, f_{n_3}^*(i))$, $q(j \mid i, f^*(i)) = \lim_{n_3\to\infty} q(j \mid i, f_{n_3}^*(i))$ for all $i, j \in S$. Replacing n_2 in (7) with n_3 and then letting $n_3 \to \infty$, we have $\rho^*\psi^*(i) = c(i, f^*(i))\psi^*(i) + \sum_{j\in S} \psi^*(j)q(j \mid i, f^*(i))$ for all $i \in S$, which implies (c_1) and also, by the extended Feymanm–Kac formula in Lemma A.1,

$$\psi^*(i) = \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{T\wedge\tau_{i_0}} (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\psi^*(x_{T\wedge\tau_{i_0}})\right] \qquad \text{for } T > 0.$$

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

Hence, using Fatou's lemma and noting $\psi^*(x_{\tau_{i_0}}) = \psi^*(i_0) = 1$, letting $T \to \infty$ we have

$$\psi^*(i) \geq \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{\tau_{i_0}} [c(x_t, f^*(x_t)) - \rho^*]\,\mathrm{d}t\right\}\right] \geq \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, \pi_t) - \rho^*)\,\mathrm{d}t\right\}\right],$$

which, together with (6), gives (c_2). □

5. The existence of R-S average optimal policies

In this section we prove the existence of a risk-sensitive average optimal stationary policy using the RS-AOE. To do so, besides Assumptions 3.1, 3.2, and 3.3, which are assumed to hold throughout this section, we need the following condition.

Assumption 5.1. $\inf_{i\in S} \psi^*(i) > 0$, *where* ψ^* *is from Theorem 4.2.*

Assumption 5.1 is new. We will show that, under Assumption 5.1 and the conditions in Theorem 4.2, a risk-sensitive average optimal stationary policy exists. Before proving this conclusion, we provide some sufficient conditions that verify Assumption 5.1.

Proposition 5.1. *Each of the following conditions implies Assumption 5.1:*

(a) The state space S is finite.

(b) Suppose that $\inf_{a\in A(i)} q(i, a) > 0$ *for all* $i \neq i_0$, *and there exists a nonnegative bounded function* V_2 *on* S *satisfying* $\sum_{j\neq i_0} q(j \mid i, a)V_2(j) \leq -1$ *for all* $a \in A(i)$, $i \neq i_0$, *and* $V_2(i_0) := 0$.

(c) (Stochastic monotonicity condition). For any $f \in F$, *c(i,f(i)) is increasing in* $i \in S$, $\sum_{j\geq k} q(j \mid i, f(i)) \leq \sum_{j\geq k} q(j \mid i+1, f(i+1))$ *for all* $i, k \in S$ *with* $k \neq i+1$, *and* $i_0 = 0$.

Proof.

(a) Take f^* and ρ^* as in Theorem 4.2. Then

$$\psi^*(i) = \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\right] > 0$$

for all $i \in S$. Thus, it is obvious that (a) implies Assumption 5.1.

(b) Let $M := \sup_{i\in S} V_2(i) < \infty$ (by the condition). Then, as in the proof of [1, Lemma 6.1.5], we have $\mathbb{E}_i^{f^*}(\tau_{i_0}) \leq V_2(i) \leq M$. Therefore, by Jensen's inequality and $c \geq 0$, we have, for all $i \in S$,

$$\psi^*(i) = \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\right] \geq \mathrm{e}^{-\rho^*\mathbb{E}_i^{f^*}(\tau_{i_0})} \geq \mathrm{e}^{-\rho^* M} \geq \mathrm{e}^{-L(i_0)M},$$

which also verifies Assumption 5.1.

(c) Obviously, it suffices to show that ψ^* is increasing on S. First, for any nonnegative increasing function u on S and $t \geq 0$, by [1, Theorem 7.3.4 and Proposition 7.3.2], we see that $\sum_{j\geq k} \mathbb{P}_i^{f^*}(x_t = j)$ is nondecreasing in $i \in S$ (for every fixed $k \in S$ at $t \geq 0$), which together with [1, Proposition 7.3.1], implies that $\mathbb{E}_i^{f^*}(u(x_t))$ is increasing in $i \in S$. For any

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

given $n \geq 1$ and $T > 0$, let $\hat{x}_k := x_{\frac{k}{2^n}T}$, $k = 0, 1, \ldots, 2^n$. Then, since $\{x_t, t \geq 0\}$ is right continuous in $t \geq 0$, by the dominated convergence theorem we have

$$\begin{aligned}\psi^*(i) &= \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\right] \\ &= \lim_{n\to\infty} \mathbb{E}_i^{f^*}\left[\exp\left\{\sum_{k=0}^{2^n} \mathbf{1}_{\{\hat{x}_k \neq 0\}}[c(\hat{x}_k, f^*(\hat{x}_k)) - \rho^*]\frac{T}{2^n}\right\}\right].\end{aligned}$$

Thus, it remains to show that

$$g(i) := \mathbb{E}_i^{f^*}\left[\exp\left\{\sum_{k=0}^{2^n} \mathbf{1}_{\{\hat{x}_k \neq 0\}}[c(\hat{x}_k, f^*(\hat{x}_k)) - \rho^*]\frac{T}{2^n}\right\}\right]$$

is increasing in $i \in S$. Fix any $i \in S$, and let $\bar{u}(j) := \exp\left\{(c(j, f^*(j)) - \rho^*)\frac{T}{2^n}\right\}$ for all $j \neq 0$ and $\bar{u}(0) := 0$. Thus, $\bar{u}(j)$ is increasing in $j \in S$, and so is $\mathbb{E}_i^{f^*}[\bar{u}(\hat{x}_{2^n})]$ in $i \in S$. Moreover, $g(i) = \mathbb{E}_i^{f^*}\left[\Pi_{k=0}^{2^n} \bar{u}(\hat{x}_k)\right] = \mathbb{E}_i^{f^*}\left[\Pi_{k=0}^{2^n-1} \bar{u}(\hat{x}_k)\mathbb{E}_{\hat{x}_{2^n-1}}^{f^*}\left[\bar{u}(\hat{x}_{2^n})\right]\right]$. Let $G^1(j) := \bar{u}(j)\mathbb{E}_j^{f^*}[\bar{u}(\hat{x}_{2^n})]$ for $j \in S$. Then, G^1 is nonnegative and increasing on S, and $g(i) = \mathbb{E}_i^{f^*}\left[\Pi_{k=0}^{2^n-2} \bar{u}(\hat{x}_k)G^1(\hat{x}_{2^n-1})\right]$. Let $G^{m+1}(j) := \bar{u}(j)G^m(j)$ for $j \in S$ and $m = 1, \ldots, 2^n - 1$. Then, by induction, we see that all G^{m+1} are nonnegative and increasing on S, and so is $g(i) = \mathbb{E}_i^{f^*}[G^{2^n}(\hat{x}_1)]$, $i \in S$. □

We now present our main result.

Theorem 5.1. *Under Assumptions 3.1, 3.2, and 3.3, and 5.1, the following hold.*

(a) There exists a solution (ρ^*, ψ^*) in $[0, L(i_0)] \times \mathbb{B}_{V_0}^+(S)$ to the RS-AOE:

$$\psi^*(i_0) = 1, \quad \rho^*\psi^*(i) = \inf_{a\in A(i)}\left\{c(i,a)\psi^*(i) + \sum_{j\in S} \psi^*(j)q(j \mid i, a)\right\} \text{ for all } i \in S, \quad (8)$$

where $\mathbb{B}_{V_0}^+(S) := \{\psi \in \mathbb{B}_{V_0}(S) : \inf_{i\in S} \psi(i) > 0\}$.

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

(c) There exists some policy f^* with $f^* \in F$ such that

(c_1) $\rho^*\psi^*(i) = c(i, f^*(i))\psi^*(i) + \sum_{j\in S} \psi^*(j)q(j \mid i, f^*(i))$ for all $i \in S$;
(c_2) for all $i \in S$,

$$\begin{aligned}\psi^*(i) &= \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\right] \\ &= \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, \pi_t) - \rho^*)\,\mathrm{d}t\right\}\right];\end{aligned}$$

(c_3) $J(i, f^*) = \rho^* = \inf_{\pi\in\Pi} J(i, \pi)$ for all $i \in S$, which means that f^* is optimal.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

Proof. Using the notation and results of Theorem 4.2, we only need to show that $\rho^* \geq J(i, f^*)$ for all $i \in S$. Indeed, for each $i \in S$, by Lemma A.1 and Assumption 5.1 that $\underline{\psi}^* := \inf_{i\in S} \psi^*(i) > 0$, we have, for all $T > 0$,

$$\psi^*(i) = \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^T (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\psi^*(x_T)\right]$$
$$\geq \underline{\psi}^*\mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^T (c(x_t, f^*(x_t)) - \rho^*)\,\mathrm{d}t\right\}\right].$$

Therefore,

$$\ln \psi^*(i) \geq \ln \underline{\psi}^* + \ln \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^T c(x_t, f^*(x_t))\,\mathrm{d}t\right\}\right] - T\rho^* \qquad \text{for } T > 0,$$

which implies that

$$\rho^* \geq \limsup_{T\to\infty} \frac{1}{T} \ln \mathbb{E}_i^{f^*}\left[\exp\left\{\int_0^T c(x_t, f^*(x_t))\,\mathrm{d}t\right\}\right] = J(i, f^*). \qquad \square$$

Remark 5.1. We have established the existence of an R-S average optimal policy in Theorem 5.1, and do not need the assumption in [7, 22, 23, 25, 27] that the state process under every stationary policy is ergodic; see Remark 7.2 for more details.

6. The computation of R-S average optimal policies

We have shown the existence of an optimal stationary policy. In this section we focus on the computational approach for finding optimal stationary policies by finite approximation.

6.1. A policy iteration algorithm

Under Assumptions 3.1–3.3 and 5.1, for each $f \in F$, by taking $A(i) := \{f(i)\}$ for all $i \in S$, it follows from Theorem 5.1 that $J(i, f)$ is independent of state i (i.e. a constant, denoted by ρ^f), which, together with the function

$$\psi^f(i) := \mathbb{E}_i^f\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f(x_t)) - \rho^f)\,\mathrm{d}t\right\}\right] \leq V_0(i) \qquad (i \in S),$$

solves the following multiplicative Poisson equation:

$$\rho\psi(i) = c(i, f(i))\psi(i) + \sum_{j\in S} \psi(j)q(j \mid i, f(i)) \qquad \text{for all } i \in S, \text{ with } \psi(i_0) = 1. \tag{9}$$

To establish the uniqueness of a solution to the Poisson equation (9) and the RS-AOE (8), we introduce the following condition.

Assumption 6.1. $\sup_{i\in S} \psi^f(i) < \infty$ *for each* $f \in F$, *where*

$$\psi^f(i) := \mathbb{E}_i^f\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f(x_t)) - \rho^f)\,\mathrm{d}t\right\}\right].$$

Since $\psi^* = \psi^{f^*}$ by Theorem 5.1(c), Assumptions 5.1 and 6.1 together with Theorem 5.1(a) imply that the ψ^* in the solution (ρ^*, ψ^*) to the RS-AOE is indeed a bounded, positive function

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

which is uniformly bounded away from zero. As mentioned in [7, Remark 5.3], showing the existence of such a solution is currently unsolved.

Obviously, Assumptions 5.1 and 6.1 are satisfied when S is finite. For the case of infinitely denumerable states, we next give suitable conditions that verify Assumptions 5.1 and 6.1.

Proposition 6.1. *Suppose that* $\hat{q} := \inf_{a\in A(i), i\neq i_0} q(i,a) > 0$, *and there exist a nonnegative bounded function* V_3 *on* S *and a constant* $\hat{\delta} > 0$ *such that* $\hat{\delta} < \hat{q}$, $V_3(i_0) = 0$, *and* $\sum_{j\neq i_0} q(j \mid i,a)V_3(j) \leq -\hat{\delta}V_3(i) - 1$ *for all* $a \in A(i)$, $i \neq i_0$. *Then, the following assertions hold.*

(a) $\psi^f(i) \geq \mathrm{e}^{-L(i_0)\sup_{i\in S} V_3(i)}$ *for all* $i \in S$ *and* $f \in F$, *which implies Assumptions 5.1.*

(b) If, in addition, $c(i,a) \leq \hat{\delta}$ *for all* $(i,a) \in K$, *then Assumptions 6.1 and 5.1 are satisfied.*

Proof. Let $M := \sup_{i\in S} V_3(i)$. Then, for any $f \in F$, it follows from the proof of [1, Lemma 6.1.5] that $\mathbb{E}_i^f(\tau_{i_0}) \leq V_3(i) \leq M$ and $\mathbb{E}_i^f\left[e^{\hat{\delta}\tau_{i_0}}\right] \leq \hat{\delta}V_3(i) + 1 \leq 1 + \hat{\delta}M < \infty$ for all $i \in S$. Thus, by

$$\psi^f(i) = \mathbb{E}_i^f\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, f(x_t)) - \rho^f)\,\mathrm{d}t\right\}\right] \geq \mathrm{e}^{-\rho^f \mathbb{E}_i^f(\tau_{i_0})} \geq \mathrm{e}^{-L(i_0)M},$$

we see that (a) is true. Obviously, (b) follows from

$$\psi^f(i) \leq \mathbb{E}_i^f\left[\exp\left\{\int_0^{\tau_{i_0}} c(x_t, f(x_t))\,\mathrm{d}t\right\}\right] \leq \mathbb{E}_i^f\left[\mathrm{e}^{\hat{\delta}\tau_{i_0}}\right] \leq 1 + \hat{\delta}M. \qquad \square$$

We next prove the uniqueness of a solution to (9) or (8).

Proposition 6.2. *Under Assumptions 3.1, 3.2, 3.3, 5.1, and 6.1, the following hold.*

(a) The solution (ρ^*, ψ^*) *to the RS-AOE (8) is unique in* $[0, L(i_0)] \times \mathbb{B}_1^+(S)$.

(b) For each $f \in F$, *the solution* (ρ^f, ψ^f) *to the multiplicative Poisson equation (9) is unique in* $[0, L(i_0)] \times \mathbb{B}_1^+(S)$.

Proof.

(a) In Theorem 5.1, we have shown that (ρ^*, ψ^*) is a solution in $[0, L(i_0)] \times \mathbb{B}_{V_0}^+(S)$ to the RS-AOE (8), and $(\rho^*, \psi^*) = (\rho^{f^*}, \psi^{f^*})$ for some $f^* \in F$. Under Assumption 6.1, it is obvious that $\psi^* \in \mathbb{B}_1^+(S)$. Hence, it remains to show that such a solution is unique to the RS-AOE (8) in $[0, L(i_0)] \times \mathbb{B}_1^+(S)$. To do so, suppose that (ρ, ψ) is an arbitrary solution in $[0, L(i_0)] \times \mathbb{B}_1^+(S)$ to the RS-AOE. Since $(\rho, \psi) \in [0, L(i_0)] \times \mathbb{B}_1^+(S)$, using a similar argument to the proofs of Theorems 4.1(b,c) and 5.1(b) yields $\rho = \inf_{\pi\in\Pi} J(i,\pi)$, and

$$\psi(i) = \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{\tau_{i_0}} (c(x_t, \pi_t) - \rho)\,\mathrm{d}t\right\}\right]$$

for all $i \in S$. It then follows that $\rho = \rho^*$ and $\psi(i) = \psi^*(i)$ for all $i \in S$.

(b) The proof is similar to part (a). $\square$

Based on the uniqueness of a solution to (9), we next provide a policy iteration algorithm for computing optimal stationary policies:

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

1. Pick an arbitrary $f \in F$. Let $k = 0$, and take $f_k := f$.
2. Policy evaluation (by Proposition 6.2(b)): Compute ρ^{f_k} and ψ^{f_k} by solving the multiplicative Poisson equation
$$\rho\psi(i) = c(i, f_k(i))\psi(i) + \sum_{j\in S} \psi(j)q(j \mid i, f_k(i)) \qquad \text{for all } i \in S, \text{ with } \psi(i_0) = 1.$$
3. Policy improvement: Obtain a policy $f_{k+1} \in F$ such that, for each $i \in S$,
$$f_{k+1}(i) := \begin{cases} f_k(i) & \text{when } B_{f_k}(i) = \emptyset, \\ a & \text{with any } a \in B_{f_k}(i) \neq \emptyset, \end{cases} \tag{10}$$
where $B_{f_k}(i) := \{a \in A(i) \mid c(i, a)\psi^{f_k}(i) + \sum_{j\in S} q(j \mid i, a)\psi^{f_k}(j) < \rho^{f_k}\psi^{f_k}(i)\}$.
4. If $f_{k+1} = f_k$ (i.e. $B_{f_k}(i) \equiv \emptyset$), then stop because f_{k+1} is optimal (by Proposition 6.2 and Theorem 5.1(c)). Otherwise, increase k by 1 and return to step 2.

6.2. A finite approximation technique

It should be noted that it is difficult to implement the policy iteration algorithm above for the case of infinite states. To compute optimal policies for infinite states, we next propose a finite approximation technique.

First, we show the convergence of this algorithm when both states and actions are finite, and the model $\mathbb{M}$ is irreducible, which means that $\{x_t, t \geq 0\}$ is irreducible under each $f \in F$.

Lemma 6.1. *Under Assumptions 3.1, 3.2, and 3.3, if S and $A(i)(i \in S)$ are finite and $\mathbb{M}$ is irreducible, then an R-S average optimal stationary policy can be obtained by the policy iteration algorithm in a finite number of steps.*

Proof. Obviously, this lemma follows from [7, Theorem 5.1] and Proposition 6.2. □

Second, in order to compute R-S average optimal policies for the case of infinite states, we construct models $\mathbb{M}_n := \{S_n, A_n(i), c_n(i, a), q_n(j \mid i, a)\}$ $(n \geq 1)$ with finite states, where $S_n := \{0, \ldots, n\}$, $A_n(i) := A(i)$, $c_n(i, a) := c(i, a)$, and
$$q_n(j \mid i, a) := \begin{cases} q(j \mid i, a) + \frac{1}{n}\sum_{k>n} q(k \mid i, a) & \text{for } 0 \leq j \leq n,\ j \neq i,\ 0 \leq i \leq n, \\ q(i \mid i, a) & \text{for } j = i \end{cases} \tag{11}$$
for each $i \in S_n$ and $a \in A_n(i)$.

Obviously, the transition rates $q_n(j \mid i, a)$ $(n \geq 1)$ are also conservative and stable. Moreover, if the function V_0 in Assumption 3.1 is nondecreasing on S, then Assumption 3.1 holds for each model $\mathbb{M}_n$, which is verified as follows. For each $n \geq 1$, $i \in S_n$, and $a \in A_n(i)$,
$$\begin{aligned} \sum_{j\in S_n} q_n(j \mid i, a)V_0(j) &= \sum_{j\in S_n, j\neq i} \left[q(j \mid i, a) + \frac{1}{n}\sum_{k>n} q(k \mid i, a)\right]V_0(j) + q(i \mid i, a)V_0(i) \\ &= \sum_{j\in S_n} q(j \mid i, a)V_0(j) + \frac{1}{n}\sum_{k>n} q(k \mid i, a)\left[\sum_{j\in S_n, j\neq i} V_0(j)\right] \\ &\leq \sum_{j\leq n} q(j \mid i, a)V_0(j) + \sum_{k>n} q(k \mid i, a)V_0(k) \\ &\leq -\delta(i)V_0(i) + b_0\mathbf{1}_{\{i_0\}} \qquad \text{(by Assumption 3.1).} \end{aligned} \tag{12}$$

Thus, in summary, we have the following theorem.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

Theorem 6.1. *Under Assumptions 3.1, 3.2, and 3.3, if the functions V_0 and V_1 are nondecreasing on S, then the following assertions hold for each $\mathbb{M}_n$ ($n \geq 1$).*

(a) The solution (ρ_n^*, ψ_n^*) to the RS-AOE (13) for $\mathbb{M}_n$ is unique in $[0, L(i_0)] \times \mathbb{B}_1^+(S)$:

$$\rho_n^* \psi_n^*(i) = \inf_{a \in A(i)} \left\{ c_n(i,a)\psi_n^*(i) + \sum_{j \in S_n} \psi_n^*(j) q_n(j|i,a) \right\} \text{ for all } i \in S_n, \tag{13}$$

$$\psi_n^*(i_0) = 1, \quad \rho_n^* \leq L_0(i_0), \quad \psi_n^*(i) \leq V_0(i) \qquad \text{for all } i \in S_n \text{ and } n \geq 1.$$

(b) There exists a policy f_n^* with $f_n^* \in F$ achieving the minimum in (13). Moreover, a policy f in F is optimal for $\mathbb{M}_n$ if and only if $f(i)$ achieves the minimum in (13) for all $i \in S_n$.

(c) If, in addition, $\mathbb{M}$ is irreducible and $A(i)$ is finite for each $i \in S$, then an R-S average optimal stationary policy f_n^* for $\mathbb{M}_n$ can be obtained by the policy iteration algorithm in a finite number of steps.

Proof. (a) and (b). Since S_n are finite, it follows from (12) that Assumptions 3.1, 3.2, 3.3, 5.1, and 6.1 are satisfied for each $\mathbb{M}_n$. Thus, (a) follows from Theorem 5.1 and Lemma 3.1(a), as well as (12), and (b) from Proposition 6.2.

(c) First, we show that $\mathbb{M}_n$ is also irreducible for each $n \geq 1$. Indeed, given any stationary policy $f_n \in F$ for model $\mathbb{M}_n$ and $i, j \in S_n$, $i \neq j$, we extend f_n to $f \in F$ by letting $f(i) := f_n(i)$ for all $i \in S_n$ and $f(i) := a_i$ for any $i \notin S_n$, where $a_i \in A(i)$ is any fixed action. Then, since $\mathbb{M}$ is irreducible, there exist $N \geq 0$ states $i_k \in S \setminus \{j\}$ ($k = 0, \ldots, N$, with $i_0 = i$, $i_{N+1} = j$) such that $q(i_{k+1} \mid i_k, f(i_k)) > 0$ for all $k = 0, \ldots, N$. If $i_k \in S_n$ for $k = 1, \ldots N$, since $i_0, i_{N+1} \in S_n$, we have $i_k \in S_n$ for all $0 \leq k \leq N+1$ and $q_n(i_{k+1} \mid i_k, f_n(i_k)) = q(i_{k+1} \mid i_k, f(i_k)) > 0$ for all $k = 0, \ldots N$, which means that i can reach j under f_n for the model $\mathbb{M}_n$. Otherwise, let $k^* := \min\{k : i_k \notin S_n\}$. Since $i_0, i_{N+1} \in S_n$, we have $1 \leq k^* \leq N$ and $\{i_0, \ldots, i_{k^*-1}\} \subset S_n$, but $i_{k^*} \notin S_n$. Thus, by the definition of the transition rates $q_n(\cdot \mid \cdot, \cdot)$ in (11) and $i_{k^*-1} \neq j \in S_n$, we have

$$q_n(j \mid i_{k^*-1}, f_n(i_{k^*-1})) = q_n(j \mid i_{k^*-1}, f(i_{k^*-1})) \geq \frac{1}{n} q(i_{k^*} \mid i_{k^*-1}, f(i_{k^*-1})) > 0,$$

which, together with $q(i_{k+1} \mid i_k, f_n(i_k)) > 0$ for $k = 0, \ldots, k^* - 1$, implies that i can reach j under f_n for the model $\mathbb{M}_n$. Thus, $\mathbb{M}_n$ is irreducible, and so (c) follows from Lemma 6.1. □

Since the sets $A(i)$ are compact, so is F. Thus, any sequence $\{f_n^*\}$ in Theorem 6.1(c) has a limit policy (say $\hat{f}^*$) in F; that is, there is a subsequence $\{f_{n_k}^*\}$ of $\{f_n^*\}$ such that $\lim_{k \to \infty} f_{n_k}^*(i) = \hat{f}^*(i)$ for each $i \in S$.

Theorem 6.2. *(Finite approximation.) Suppose that Assumptions 3.1, 3.2, and 3.3 are satisfied, and that V_0 is nondecreasing on S. Then the following hold.*

(a) A sequence $\{f_n^*\}$ of R-S average optimal stationary policies f_n^* for $\mathbb{M}_n$ exists, and it has a limit policy $\hat{f}^*$ such that, for all $i \in S$,

$$\hat{\rho}\hat{\psi}(i) = c(i, \hat{f}^*(i))\hat{\psi}(i) + \sum_{j \in S} q(j \mid i, \hat{f}^*(i))\hat{\psi}(j)$$

$$= \inf_{a \in A(i)} \left\{ c(i,a)\hat{\psi}(i) + \sum_{j \in S} q(j \mid i, a)\hat{\psi}(j) \right\} \tag{14}$$

for some function $\hat{\psi}$ on S such that $\hat{\psi} \leq V_0$.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

(b) If, in addition, the condition in Proposition 6.1 holds with a nondecreasing V_3 on S and a constant $\hat{\delta} \geq c(i,a)$ on K, then the policy $\hat{f}^*$ in (a) is optimal for the model $\mathbb{M}$.

Proof.

(a) Suppose that $\hat{f}^*(i) = \lim_{k\to\infty} f^*_{n_k}(i)$ for all $i \in S$. Then, by Theorem 6.1, we have $0 \leq \rho^*_{n_k} \leq L_0(i_0)$ and $\psi^*_{n_k}(i) \leq V_0(i)$ for all $k \geq 1$. Thus, diagonalization arguments ensure the existence of a subsequence $\{n_{k_l}, l \geq 1\}$ of $\{n_k, k \geq 1\}$ and $(\hat{\rho}, \hat{\psi}) \in [0, L_0(i_0)] \times \mathbb{B}_{V_0}(S)$ such that

$$\lim_{l\to\infty} f^*_{n_{k_l}}(i) = \hat{f}^*(i), \quad \lim_{l\to\infty} \rho^*_{n_{k_l}} = \hat{\rho}, \quad \lim_{l\to\infty} \psi^*_{n_{k_l}}(i) = \hat{\psi}(i) \leq V_0(i)$$
$$\text{for all } i \in S \text{ with } \hat{\psi}(i_0) = 1. \quad (15)$$

For given $i \in S$, there exists $\tilde{n}$ such that $i \in S_n$ for $n \geq \tilde{n}$. Then, Theorem 6.1(a,b) give

$$\begin{aligned}\rho^*_{n_{k_l}}\psi^*_{n_{k_l}}(i) &= c_{n_{k_l}}(i, f^*_{n_{k_l}}(i))\psi^*_{n_{k_l}}(i) + \sum_{j\in S_{n_{k_l}}} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j \mid i, f^*_{n_{k_l}}(i)) \text{ for all } l \geq \tilde{n} \\ &= \inf_{a\in A(i)} \left\{ c_{n_{k_l}}(i,a)\psi^*_{n_{k_l}}(i) + \sum_{j\in S_n} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j \mid i, a) \right\} \\ &\leq c_{n_{k_l}}(i,a)\psi^*_{n_{k_l}}(i) + \sum_{j\in S_n} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j \mid i, a) \qquad \text{for all } a \in A(i). \quad (16)\end{aligned}$$

On the other hand, since $\psi^*_n \leq V_0$ for all $n \geq 1$, by (11) and (15) we have

$$\sum_{j\in S_n} \psi^*_n(j) q_n(j \mid i, f^*_n(i)) = \sum_{j\in S_n} \psi^*_n(j) q(j \mid i, f^*_n(i)) + \frac{1}{n} \sum_{k>n} q(k \mid i, f^*_n(i)) \sum_{0\leq j\leq n, j\neq i} \psi^*_n(j), \quad (17)$$

which, together with the monotonicity of V_0 and

$$0 \leq \frac{1}{n} \sum_{k>n} q(k \mid i, f^*_n(i)) \sum_{0\leq j\leq n, j\neq i} \psi^*_n(j) \leq \sum_{k>n} q(k \mid i, f^*_n(i)) V_0(k) \to 0 \text{ as } n \to \infty,$$

implies that

$$\lim_{l\to\infty} \left[\sum_{j\in S_{n_{k_l}}} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j \mid i, f^*_{n_{k_l}}(i)) \right] = \sum_{j\in S} \hat{\psi}(j) q(j \mid i, \hat{f}(i)). \quad (18)$$

Thus, by (15)–(18), we get (14).

(b) As the proof of (12), we have $\hat{\delta} < q_n(i,a)$ and $\sum_{j\neq i_0} q_n(j \mid i, a) V_3(j) \leq -\hat{\delta} V_3(i) - 1$ for all $a \in A(i)$ and every $i \neq i_0$, $n \geq 1$, which, together with the proofs of Proposition 6.1 and Theorem 5.1(c), give $e^{-L(i_0)M} \leq \psi^*_n(i) \leq 1 + \hat{\delta} M$, where $M = \sup_{i\in S} V_3(i) < \infty$, and so $e^{-L(i_0)M} \leq \hat{\psi}^*(j) \leq 1 + \hat{\delta} M$ for all $i \in S$. Hence, by (14) and Proposition 6.2, we see that (b) is also true. □

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

7. Example

In this section we provide an example to illustrate the conditions in this paper and show the effectiveness of the finite approximation in Theorem 6.2 for risk-sensitive average CTMDPs.

Example 7.1. (*Controlled population processes.*) In a population process, we regard the population size at any time as the state variable, and suppose that the birth and death parameters can be controlled by a decision maker; they are denoted by $\lambda_i(a_1)$ and $\mu_i(a_2)$ respectively, and may depend on the system's state i and decision variables (a_1, a_2) taken by the decision maker. When the state of the process is at $i \in S := \{0, 1, \dots, \}$, the decision maker takes an action $a := (a_1, a_2)$ from a given set $A(i) =: A_1(i) \times A_2(i)$, which may increase or decrease the parameters $\lambda_i(a_1)$ and $\mu_i(a_2)$. On the other hand, because of some possible catastrophe, it is suitable to suppose that a transition from i to 0 may happen at rate $\beta(i)$ for all $i \geq 1$. Choosing any action $a = (a_1, a_2)$ at state i results in some cost, denoted by $c(i, a)$. Moreover, the decision maker wishes to minimize the associated risk-sensitive average cost.

We now formulate the controlled population processes as CTMDPs. Obviously, the state space $S = \{0, 1, \dots\}$ is denumerable; the corresponding transition rates $q(j \mid i, a)$ are as follows. When there is no population in the system (i.e. $i = 0$), any control of death is unnecessary, and so we set $A_2(0) := \{0\}$. Thus, we have $q(1 \mid 0, a) = -q(0 \mid 0, a) := \lambda_0(a_1)$ for $a = (a_1, a_2) \in A_1(0) \times A_2(0)$, where a_1 denotes immigration rates varying in $A_1(0) := [0, \alpha_1]$ for some constant $\alpha_1 > 0$. Moreover, for each $i \geq 1$, $a = (a_1, a_2) \in A_1(i) \times A_2(i)$, we have

$$q(j \mid i, a) = \begin{cases} \lambda_i(a_1) & \text{if } j = i + 1, \\ -\lambda_i(a_1) - \mu_i(a_2) - \beta(i) & \text{if } j = i, \\ \mu_i(a_2) & \text{if } j = i - 1,\ i \geq 2, \\ \mu_1(a_2) + \beta(1) & \text{if } j = 0,\ i = 1, \\ \beta(i) & \text{if } j = 0,\ i \geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

We aim to find conditions imposed on $q(j \mid i, a)$ in (19) and $c(i, a)$ that can ensure the existence of an optimal policy, and thus consider the following sets of hypotheses H_1 and H_2 with given positive constants μ and λ. A characterization of these hypotheses is that their conditions are imposed on the elements of the model and thus can be verified.

H_1 (On controlled birth and death processes with catastrophes [1, p. 292]):

(a) $A(i) := [-\lambda, \lambda] \times [-\mu, \mu]$ for all $i \geq 1$ and $\mu \geq \max\{\lambda, \frac{1}{2}\}$;

(b) $\lambda_i(a_1) := \lambda i + a_1$ for all $i \geq 0$, and $\mu_i(a_2) =: \mu(i + 2)^2 + a_2$ for all $i \geq 1$;

(c) $\beta_* := \inf_{i \geq 1} \beta(i) > 0$;

(d) $0 \leq c(i, a) \leq \ln \sqrt{i + 2}$ for all $i \geq 0$ and $a \in A(i)$, and $c(i, a)$ is continuous in $a \in A(i)$ for each fixed $i \in S$.

Remark 7.1. The transition and cost rates in the condition H_1 can be unbounded. However, H_1 needs the catastrophe hypothesis (i.e. $\inf_{i \geq 1} \beta(i) > 0$), which is for the validity of Proposition 5.1(b). To remove the catastrophe hypothesis, we need stochastic monotonicity, and thus modify H_1 as the following H_2, which is for the verification of the conditions in Proposition 5.1(c).

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

H_2 (On controlled birth and death processes without any catastrophe [1, p. 292]):

(a) $A_1(0) = [0, \alpha_1]$, $A(i) := [-\lambda, \lambda] \times [\mu, \mu]$ for all $i \geq 1$, where $\mu \geq \max\{\lambda, \frac{1}{2}\}$;

(b) $\lambda_i(a_1) := \lambda i + a_1$ for $i \geq 0$, and $\mu_i(a_2) =: \mu(i+2)^2 + a_2$ for all $i \geq 1$;

(c) $\beta(i) = 0$ for all $i \geq 1$;

(d) $0 \leq c(i, a) \leq \ln \sqrt{i+2}$ for all $(i, a) \in K$, and $c(i, a)$ is continuous in $a \in A(i)$ for each fixed $i \in S$;

(e) $\inf_{a' \in A(i+1)} c(i+1, a') \geq c(i, a)$ for all $(i, a) \in K$, which implies that $c(i, f(i))$ is increasing in $i \in S$ for any given $f \in F$.

Proposition 7.1. *Under one of H_1 and H_2, Assumptions 3.1, 3.2, 3.3, and 5.1 are satisfied. Thus, an optimal stationary policy exists for Example 7.1.*

Proof. Under H_1, in order to verify the assumptions, let $i_0 := 0$, $V_0(i) := (i+2)^2$, $\delta(i) := \ln \sqrt{i+2}$, $V_1(i) := (i+2)^4$, $V_2(i) := \frac{1}{\beta_*}$ for $i \neq 0$, $V_2(0) := 0$. Then, using the conditions in H_1, a direct calculation gives, for all $i \geq 1$ and $(a_1, a_2) \in A(i)$,

$$\sum_{j \in S} q(j \mid 0, a) V_0(j) = 5a_1 \leq -\delta(0) V_0(0) + 8\mu + 5b_1, \tag{20}$$

$$\begin{aligned} \sum_{j \in S} q(j \mid i, a) V_0(j) &= -[\mu(i+2)^2 + a_2](2i+3) + (\lambda i + a_1)(2i+5) - \beta(i)(2i^2+4) \\ &\leq -\mu(i+2) V_0(i) - \mu(i+2)^2(i+1) + \mu(2i+3) + \lambda(i+1)(2i+5) \\ &\leq -\frac{\mu}{2}(i+2) V_0(i) \leq -\delta(i) V_0(i), \end{aligned} \tag{21}$$

$$\begin{aligned} \sum_{j \in S} q(j \mid i, a) V_1^2(j) &= [a_2 + (i+2)^2 \mu][(i+1)^8 - (i+2)^8] + (a_1 + \lambda i)[(i+3)^8 - (i+2)^8] \\ &\quad + \beta(i)[2^8 - (i+2)^8] \\ &\leq 192(\lambda + \mu)(i+2)^8 = 192(\lambda + \mu) V_1^2(i), \end{aligned} \tag{22}$$

$$\sum_{j \neq 0} q(j \mid i, a) V_2(j) = -\frac{\beta(i)}{\beta_*} \leq -1. \tag{23}$$

Thus, since $q^*(i) \leq (\lambda + \mu)(i+2)^2 \leq (\lambda + \mu) V_0(i)$, Assumption 3.1 follows from (20) and (21). From the descriptions of the example and H_1, we see that Assumption 3.3 is satisfied, and (22) implies Assumption 3.2. It remains to verify Assumption 5.1. Indeed, since $\inf_{a \in A(i)} q(i, a) = \inf_{a \in A(i)} (\lambda i + a_1 + \mu(i+2)^2 + a_2 + \beta(i)) \geq 3\mu > 0$ (by H_1) for each $i \neq 0$, by (23) and Proposition 5.1(b) we see that Assumption 5.1 is satisfied.

Under H_2, as for the arguments for H_1, we see that Assumptions 3.1–3.3 are all satisfied. Moreover, since the birth and death processes without any catastrophe (by H_2(c)) are stochastic monotone and the second part of H_2(d) implies that $c(i, f(i))$ is nondecreasing in $i \in S$ (for any given stationary policy f), Assumption 5.1 follows from Proposition 5.1(c). Thus, Theorem 5.1 gives the desired result. □

Remark 7.2. Under H_1 or H_2, take any stationary policy f with $f(0) = (0, 0)$ (or $f(1) = (-\mu, -\lambda)$). Then, we have $q(0 \mid 0, f(0)) = 0$ (or $q(1 \mid 1, f(1)) = 0$), and so the state 0 (or 1) is absorbing under the policy f, while the state process under every stationary policy has been

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

assumed to be ergodic in [7, 22, 23, 25, 27]. Of course, if the sets $A(0)$ and $A(i)$ $(i \geq 1)$ are taken to be compact subsets of $(0, b_1]$ and $(-\lambda, \lambda) \times (-\mu, \mu)$ respectively, then $\mathbb{M}$ (i.e. process $\{x_t, t \geq 0\}$) is irreducible under any stationary policy.

Remark 7.3. In the conditions H_1 and H_2 for Example 7.1, the cost and transition rates are allowed to be unbounded. Moreover, if the $c(i, a)$ in $H_1(d)$ is unbounded (for example, $c(i, a) := \sqrt{\ln\left(i + 1 + \frac{|a_1|+|a_2|}{\mu+\lambda}\right)}$, then the smallness condition (stronger than the standard boundedness condition) on the costs in [7, 23] is not satisfied. If taking $c(i, a) = \sqrt{\ln\left[1 + \frac{|a_1|+|a_2|}{(\mu+\lambda)(i+1)}\right]}$ which satisfied $H_1(d)$, then $\liminf_{i\to\infty} \inf_{a\in A(i)} c(i, a) = 0$. This shows that the near-monotone condition in [22] (i.e. $\liminf_{i\to\infty} \inf_{a\in A(i)} c(i, a) > \inf_{f\in F} J(i, f)$) fails to hold. Furthermore, the uniform boundedness hypothesis on the transition rates in [7, 22, 23, 25] fail to hold for $H_1(b)$.

To illustrate the calculation of optimal policies for the case of infinite states by the finite approximation policy iteration algorithm, we consider the following conditions. (On controlled irreducible birth and death processes with catastrophes [1, p. 292])):

(a) $A_1(0) := [\alpha_2, \alpha_3]$ for constants $\alpha_3 > \alpha_2 > 0$, and $A(i) := [0, \lambda] \times [0, \mu]$ for $i \geq 1$;

(b) $\lambda_i(a_1) := \lambda i + a_1$ for all $i \geq 0$, $\mu_i(a_2) =: \mu i + a_2$ for all $i \geq 1$, and $\mu \geq \lambda$;

(c) $\beta(i) \geq \max\{2 + 3(\mu + \lambda) + L, 4\sqrt{\ln(1 + i)} + 2\lambda - \mu\}$ for $i \geq 1$, with a constant $L > 0$;

(d) $0 \leq c(i, a) \leq \min\{L, \frac{1}{4}(2 + 4\mu + \lambda + L)\}$ for all $i \geq 0$ and $a \in A(i)$, and $c(i, a)$ is continuous in $a \in A(i)$ for each fixed $i \in S$.

Proposition 7.2. *Under H_3 for Example 7.1, an optimal policy exists and can be obtained as a limit policy of $\{f_n^*\}$ of optimal policies f_n^* for the models $\mathbb{M}_n$.*

Proof. Let $i_0 := 0$, $V_0(i) := 1 + i$, $\delta(i) := \frac{1}{4}(\beta(i) + \mu - 2\lambda)$, $V_1(i) := (1 + i)^2$, $V_3(i) := \frac{i}{1+i}$ for all $i \geq 0$, and $\hat{\delta} := L$. Then V_0 and V_3 are nondecreasing on S, and $q_* = \inf_{a\in A(i), i\geq 1} q(i, a) = \lambda + \mu > \hat{\delta}$. Moreover, simple calculations give $c(i, a) \leq \delta(i) \leq \sqrt{\ln V_0(i)} (a \in A(i))$ and

$$\sum_{j\in S} q(j \mid i, a) V_0(j) \leq -\delta(i) V_0(i) + 3\alpha_3 \mathbf{1}_{\{0\}}(i) \quad \text{for } i \in S,$$

$$\sum_{j\in S} q(j \mid i, a) V_1^2(j) = p_1 + p_2 i + p_3 i^2 + p_4 i^3 + p_4 i^4 \leq K_1 V_1^2(i) + K_2 \quad \text{for } i \in S,$$

$$\sum_{j\in S} q(j \mid 0, a) V_3(j) = \frac{a_1}{2} \leq \alpha_3,$$

$$\sum_{j\neq 0} q(j \mid i, a) V_3(j) \leq -L V_3(i) - 1 \quad \text{for } i \geq 1,$$

where the constants $p_1, p_2, p_3, p_4, K_1, K_2$ are determined by the given λ, μ, and $\beta(i)$. Thus, from the inequalities above and Proposition 6.1 we see that all the conditions in Theorem 6.2(b) are satisfied, and thus the result follows from Theorem 6.2. □

To show the finite approximation policy iteration algorithm, we conduct a numerical experiment under H_3. To do so, we take some specific data for the model. In particular, we take $\alpha_2 = 0.7, \alpha_3 = 0.9, \lambda = 0.5, \mu = 0.5, L = 2, A_1(0) = \{\alpha_2, \alpha_3\}, A(i) = \{0, \lambda\} \times \{0, \mu\}$ for $i \geq 1, \beta(i) =$

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

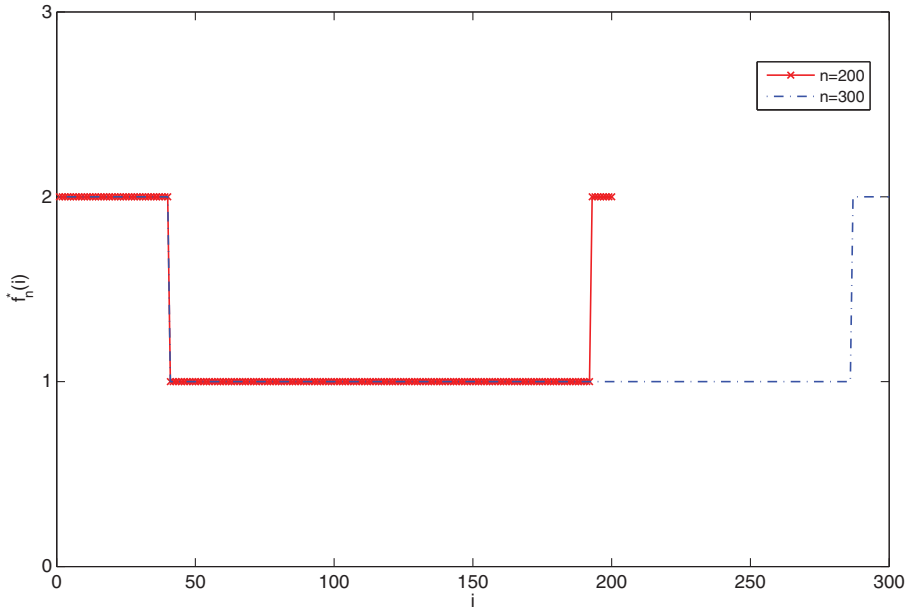


FIGURE 1: Optimal policies.

$\max\{2+3(\mu+\lambda)+L, 4\sqrt{\ln(1+i)}+2\lambda-\mu\}$ for $i\geq 1$, $c(0,\alpha_2)=0.5$, $c(0,\alpha_3)=0.8$, and $c(i,\{0\}\times\{0\})=c(i,\{0\}\times\{\mu\})=0.4+i/(i+5)$, $c(i,\{\lambda\}\times\{0\})=0.45+i/(i+5)$, $c(i,\{\lambda\}\times\{\mu\})=0.5+i/(i+5)$ for $i\geq 1$. We implement the finite approximation policy iteration algorithm with these data in Matlab. Our numerical algorithm can run for every n. However, it costs time for large n. We found that $\rho_n^*=0.5158$ for $n=10$, 20, 30, 50, 100, 200, 300. The optimal policies f_n^* for some n are listed below.

$n=10$: $f_n^*(0)=\alpha_2, f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,9, f_n^*(i)=\{0\}\times\{0\}$ for $i=10$.

$n=20$: $f_n^*(0)=\alpha_2, f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,17, f_n^*(i)=\{0\}\times\{0\}$ for $i=18, 19, 20$.

$n=30$: $f_n^*(0)=\alpha_2$, $f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,25$, $f_n^*(i)=\{0\}\times\{0\}$ for $i=26,\ldots,30$.

$n=50$: $f_n^*(0)=\alpha_2$, $f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,39$, $f_n^*(i)=\{0\}\times\{0\}$ for $i=40,\ldots,50$.

$n=100$: $f_n^*(0)=\alpha_2$, $f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,40$, $f_n^*(i)=\{0\}\times\{0\}$ for $i=41,\ldots,100$.

$n=200$: $f_n^*(0)=\alpha_2$, $f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,40$, $f_n^*(i)=\{0\}\times\{0\}$ for $i=41,\ldots,192, f_n^*(i)=\{0\}\times\{\mu\}$ for $i=193,\ldots,200$.

$n=300$: $f_n^*(0)=\alpha_2$, $f_n^*(i)=\{0\}\times\{\mu\}$ for $i=1,\ldots,40$, $f_n^*(i)=\{0\}\times\{0\}$ for $i=41,\ldots,286, f_n^*(i)=\{0\}\times\{\mu\}$ for $i=287,\ldots,300$.

To see the numerical results more clearly, we plot the optimal policies f_n^* for $n=200$ and $n=300$ in Figure 1, where the symbols 1,2,3,4 represent the actions $\{0\}\times\{0\}$, $\{0\}\times\{\mu\}$, $\{\lambda\}\times\{0\}$, $\{\lambda\}\times\{\mu\}$, respectively.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

Appendix A. Proofs

In this appendix we provide the proofs of Theorems 3.1 and 4.1, as well as Lemmas 3.1 and 3.2, which require the following lemma.

Lemma A.1. *(Extension of Feynman–Kac formula.) Under Assumptions 3.1 and 3.2, for any $T > 0$ the following assertions hold, where $c(i, \pi_v) := \int_{A(i)} c(i, a)\pi_v(\mathrm{d}a \mid i)$ and $q(j \mid i, \pi_t) := \int_{A(i)} q(j \mid i, a)\pi_t(\mathrm{d}a \mid i)$ for $i, j \in S$, $t \geq 0$.*

(a) For any $\pi \in \Pi$ and $u \in \mathbb{B}^{ac}_{V_0,V_1}([0, T] \times S)$,

$$\begin{aligned}
&\mathbb{E}^{\pi}_{\gamma}\Bigg[\int_0^{T\wedge\tau_D}\Bigg(\frac{\partial\Big(\exp\Big\{\int_0^t\int_A c(x_v, a)\pi(\mathrm{d}a \mid \omega, v)\,\mathrm{d}v\Big\}u(t, x_t)\Big)}{\partial t}\\
&\qquad + \sum_{j\in S}\exp\Big\{\int_0^t\int_A c(x_v, a)\pi(\mathrm{d}a \mid \omega, v)\,\mathrm{d}v\Big\}u(t, j)\\
&\qquad\qquad \times\int_A q(j \mid x_t, a)\pi(\mathrm{d}a \mid \omega, t)\Bigg)\mathrm{d}t \mid x_0 = i\Bigg]\\
&= \mathbb{E}^{\pi}_{\gamma}\Bigg[\exp\Big\{\int_0^{T\wedge\tau_D}\int_A c(x_v, a)\pi(\mathrm{d}a \mid \omega, v)\,\mathrm{d}v\Big\}u(T\wedge\tau_D, x_{T\wedge\tau_D}) \mid x_0 = i\Bigg] - u(0, i),
\end{aligned}$$

$i \in S$, $D \subseteq S$, where $\{x_t, t \geq 0\}$ may be not Markovian since the policy π may depend on histories.

(b) For each $\pi = \pi_t(\mathrm{d}a \mid \cdot) \in \Pi^{\mathrm{r}}_m$ and $u \in \mathbb{B}^{ac}_{V_0,V_1}([0, T] \times S)$,

$$\begin{aligned}
&\mathbb{E}^{\pi}_{\gamma}\Bigg[\int_s^{T\wedge\tau_D}\Bigg(\frac{\partial\Big(\exp\Big\{\int_s^t c(x_v, \pi_v)\,\mathrm{d}v\Big\}u(t, x_t)\Big)}{\partial t}\\
&\qquad + \sum_{j\in S}\Big(\exp\Big\{\int_s^t c(x_v, \pi_v)\,\mathrm{d}v\Big\}u(t, j)\Big)q(j \mid x_t, \pi_t)\Bigg)\mathrm{d}t \mid x_s = i\Bigg]\\
&= \mathbb{E}^{\pi}_{\gamma}\Bigg[\exp\Big\{\int_s^{T\wedge\tau_D} c(x_v, \pi_v)\,\mathrm{d}v\Big\}u(T\wedge\tau_D, x_{T\wedge\tau_D}) \mid x_s = i\Bigg] - u(s, i)
\end{aligned}$$

for all $(s, i) \in [0, T] \times S$, $D \subseteq S$.

Proof. By Assumption 3.1(iii), we have $2tc(i, a) \leq 2t\delta(i) \leq 2t\sqrt{\ln V_0(i)} \leq t^2 + \ln V_0(i)$ for all $t \geq 0$, and thus $\mathrm{e}^{2Tc(i,a)} \leq \mathrm{e}^{T^2}V_0(i)$ for all $i \in S$, which verifies Assumption 3.1(iii) with $M_0 := \mathrm{e}^{T^2}$ in [9]. Thus, from the proof of [9, Theorem 3.1], replacing T with $T \wedge \tau_D$, we see that this lemma is also true. □

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

A.1. Proof of Lemma 3.1

(a) Obviously, the results hold for $i = i_0$. For any $i \neq i_0$, since $V_0 \geq 1$, by Lemma A.1(a) for $u := V_0$ and $c(i, \cdot) := \delta(i)$ we have

$$\begin{aligned}\mathbb{E}_i^\pi\left[\exp\left\{\int_0^{T\wedge\tau_{i_0}}\delta(x_v)\,\mathrm{d}v\right\}\right] - V_0(i) &\leq \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{T\wedge\tau_{i_0}}\delta(x_v)\,\mathrm{d}v\right\}V_0\big(x_{T\wedge\tau_{i_0}}\big)\right] - V_0(i)\\ &= \mathbb{E}_i^\pi\left[\int_0^{T\wedge\tau_{i_0}}\exp\left\{\int_0^t\delta(x_v)\,\mathrm{d}v\right\}\left(\delta(x_t)V_0(x_t) + \sum_{j\in S}V_0(j)q(j\mid x_t,\pi_t)\right)\mathrm{d}t\right]\\ &\leq b_0\mathbb{E}_i^\pi\left[\int_0^{T\wedge\tau_{i_0}}\exp\left\{\int_0^t\delta(x_v)\,\mathrm{d}v\right\}\mathbf{1}_{\{i_0\}}(x_t)\,\mathrm{d}t\right] = 0,\end{aligned}$$

which, together with letting $T \to \infty$, gives the desired result.

(b) Since $\delta_* = \inf_{i\neq i_0}\delta(i) > 0$, by (a) and the Jensen inequality, we have

$$\delta_*\mathbb{E}_i^\pi[\tau_{i_0}] \leq \mathrm{e}^{\delta_*\mathbb{E}_i^\pi[\tau_{i_0}]} \leq \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{\tau_{i_0}}\delta(x_v)\,\mathrm{d}v\right\}\right] \leq V_0(i),$$

which implies (b).

A.2 Proof of Theorem 3.1

(a) Since Assumption 3.1(iii) in [9] has been verified in the proof of Lemma A.1, for any $i \in S$, by [9, Theorem 4.1] and an obvious change of time, we have

$$\begin{cases}\varphi'(t,i) = \inf_{a\in A(i)}\left[c(i,a)\varphi(t,i) + \sum_{j\in S}\varphi(t,j)q(j\mid i,a)\right] \text{ almost surely for } t\in[0,T],\\ \varphi(0,i) = 1.\end{cases}$$

Furthermore, for any $t_1, t_2 \in [0, T]$, $\varepsilon > 0$, Assumption 3.3 together with [15, Lemma 4.1] ensures the existence of $a(i, t_1, t_2) \in A(i)$ and a finite subset $S(i)$ of S (depending on i) such that $i \in S(i)$ and $\sup_{a\in A(i)}\sum_{j\notin S(i)}|\varphi(t_1,j) - \varphi(t_2,j)||q(j\mid i,a)| = \sum_{j\notin S(i)}|\varphi(t_1,j) - \varphi(t_2,j)||q(j\mid i,a(i,t_1,t_2))|$, and $\sum_{j\notin S(i)}V_0(j)q(j\mid i,a) < \varepsilon$ for all $a \in A(i)$. Thus, for any $t_1, t_2 \in [0, T]$, we have

$$\begin{aligned}&\|c(i,\cdot)\||\varphi(t_1,i) - \varphi(t_2,i)| + \sup_{a\in A(i)}\sum_{j\in S}|\varphi(t_1,j) - \varphi(t_2,j)||q(j\mid i,a)|\\ &\quad\leq \|c(i,\cdot)\||\varphi(t_1,i) - \varphi(t_2,i)| + \sum_{j\in S(i)}|\varphi(t_1,j) - \varphi(t_2,j)|q^*(i)\\ &\qquad + 2\|\varphi\|_{V_0}\sum_{j\notin S(i)}V_0(j)q(j\mid i,a(i,t_1,t_2))\\ &\quad\leq \|c(i,\cdot)\||\varphi(t_1,i) - \varphi(t_2,i)| + \sum_{j\in S(i)}|\varphi(t_1,j) - \varphi(t_2,j)|q^*(i) + 2\|\varphi\|_{V_0}\varepsilon,\end{aligned}$$

which, together with the continuity of $\varphi(t, i)$ in t and the finiteness of $\|c(i, \cdot)\|$, $S(i)$ and $q^*(i)$, implies that $\inf_{a\in A(i)}[c(i, a)\varphi(t, i) + \sum_{j\in S}\varphi(t, j)q(j\mid i, a)]$ is continuous in $t \geq 0$ for any fixed $i \in S$. Hence, by the first equation in this proof, $\varphi'(t, i)$ exist for all $t \geq 0$ and $\varphi'(t, i)$ is continuous in $t \geq 0$.

(b) follows from Theorem 4.2(c) in [9].

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

(c) Let $\pi = f_t$ be any Markov policy in $\Pi_{\mathrm{m}}^{\mathrm{d}}$, and $\pi^* = f_t^*$ the deterministic Markov policy from part (b). Define a policy $\hat{\pi}$ by $\hat{\pi}(\mathrm{d}a \mid \omega, s) = \mathbf{1}_{\{t\wedge\tau_D > s\}}\delta_{f_s(x_s(\omega))}(\mathrm{d}a) + \mathbf{1}_{\{t\wedge\tau_D \le s\}}\delta_{f_s^*(x_s(\omega))}(\mathrm{d}a)$. Let $\mathcal{F}_{t\wedge\tau_D}$ be the σ-algebra generated by the stopping time $t\wedge\tau_D$ with respective the filtration $\mathcal{F}_s := \sigma(x_v,\, v \le s)$. Since $\varphi(t, i) = \varphi(t, i, \pi^*)$ for $t \in [0, T]$ and $i \in S$, we have

$$
\begin{aligned}
\varphi(t, i) &\le \mathbb{E}_i^{\hat{\pi}}\left[\exp\left\{\int_0^t\int_A c(x_s, a)\hat{\pi}(\mathrm{d}a \mid \omega, s)\,\mathrm{d}s\right\}\right] \\
&= \mathbb{E}_i^{\hat{\pi}}\left[\exp\left\{\int_0^{t\wedge\tau_D} c(x_s, f_s(x_s))\,\mathrm{d}s\right\}\mathbb{E}_i^{\hat{\pi}}\left[\exp\left\{\int_{t\wedge\tau_D}^t c(x_s, f_s(x_s))\,\mathrm{d}s\right\} \mid \mathcal{F}_{t\wedge\tau_D}\right]\right] \\
&= \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_D} c(x_s, f_s(x_s))\,\mathrm{d}s\right\}\varphi(t - t\wedge\tau_D, x_{t\wedge\tau_D}, \pi^*)\right] \\
&= \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_D} c(x_s, f_s(x_s))\,\mathrm{d}s\right\}\varphi(t - t\wedge\tau_D, x_{t\wedge\tau_D})\right] \\
&\le \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_D} c(x_s, f_s(x_s))\,\mathrm{d}s\right\}\varphi(t - t\wedge\tau_D, x_{t\wedge\tau_D}, \pi)\right] = \varphi(t, i, \pi).
\end{aligned}
$$

Taking the infimum over $\pi \in \Pi_{\mathrm{m}}^{\mathrm{d}}$ on both sides of the above inequality and using (b) again, we see that part (c) holds.

A.3 Proof of Lemma 3.2

(a) Since $1 = \varphi(0, i) \le \varphi(t, i)$ for all $i \in S$ and $t \ge 0$, by Theorem 3.1(c) we have, for each $\pi \in \Pi$,

$$
\begin{aligned}
\varphi(t, i) &\le \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t - t\wedge\tau_{i_0}, x_{t\wedge\tau_{i_0}})\right] \\
&= \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t - t\wedge\tau_{i_0}, x_{t\wedge\tau_{i_0}})\mathbf{1}_{\{t\le\tau_{i_0}\}}\right] \\
&\quad + \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t - t\wedge\tau_{i_0}, x_{t\wedge\tau_{i_0}})\mathbf{1}_{\{t>\tau_{i_0}\}}\right] \\
&= \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\mathbf{1}_{\{t\le\tau_{i_0}\}}\right] \\
&\quad + \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t - \tau_{i_0}, x_{\tau_{i_0}})\mathbf{1}_{\{t>\tau_{i_0}\}}\right] \\
&\le \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t, x_{\tau_{i_0}})\mathbf{1}_{\{t\le\tau_{i_0}\}}\right] \\
&\quad + \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{t\wedge\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t, x_{\tau_{i_0}})\mathbf{1}_{\{t>\tau_{i_0}\}}\right] \\
&\le \mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi(t, x_{\tau_{i_0}})\right] \\
&= \varphi(t, i_0)\mathbb{E}_i^{\pi}\left[\exp\left\{\int_0^{\tau_{i_0}} c(x_v, \pi_v)\,\mathrm{d}v\right\}\right],
\end{aligned}
$$

which, together with Lemma 3.1(a), completes the proof of part (a).

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

(b) $\varphi(t, i)$ is increasing in t for each $i \in S$, and thus $\varphi'(t, i) \geq 0$. Moreover, for each $t \geq 0$, by (a) and Assumption 3.1(i) we have

$$\begin{aligned} 0 \leq \frac{\varphi'(t, i)}{\varphi(t, i_0)} &= \inf_{a \in A(i)} \left[c(i, a) \frac{\varphi(t, i)}{\varphi(t, i_0)} + \sum_{j \in S} \frac{\varphi(t, j)}{\varphi(t, i_0)} q(j \mid i, a) \right] \\ &\leq \inf_{a \in A(i)} \left[\delta(i) V_0(i) + \sum_{j \in S} V_0(j) |q(j \mid i, a)| \right] \\ &= \inf_{a \in A(i)} \left[\delta(i) V_0(i) + \sum_{j \in S} V_0(j) q(j \mid i, a) - 2q(i \mid i, a) V_0(i) \right] \\ &\leq 2q^*(i) V_0(i) + b_0, \end{aligned}$$

and so part (b) follows.

A.4 Proof of Theorem 4.1

As in the arguments in [23], we prove Theorem 4.1 by Theorem 3.2 and Lemma 3.2.

(a) For any fixed $n \geq 1$, and for $t \geq 0$, $i \in S$, let

$$\varphi_n(t, i) := \inf_{\pi \in \Pi_{\mathrm{m}}^{\mathrm{d}}} \mathbb{E}_i^{\pi} \left[\exp \left\{ \int_0^t c_n(x_v, \pi_v) \, \mathrm{d}v \right\} \right] (\geq 0), \quad \hat{\varphi}_n(t, i) := \frac{\varphi_n(t, i)}{\varphi_n(t, i_0)}. \tag{24}$$

Then, since $0 \leq c_n \leq c$, the conditions of Theorem 3.1 and Lemma 3.2 still hold when c is replaced with c_n. Thus, using Lemma 3.2(a) and (24), we have $0 \leq \hat{\varphi}_n(t, i) \leq V_0(i)$ for all $t \geq 0$. Thus, for any $m \geq 1$, the mean value theorem together with Theorem 3.1(a) gives the existence of $k_0(i, m) \in [m, 2m]$ (depending on the given m and i) such that

$$\hat{\varphi}_n'(k_0(m, i), i) = \frac{\hat{\varphi}_n(2m, i) - \hat{\varphi}_n(m, i)}{2m - m} \to 0 \qquad (\text{as } m \to \infty). \tag{25}$$

Since S is denumerable, diagonalization arguments as well as (25) ensure the existence of a subsequence $\{k_1\}$ of $\{k_0(i, m), m \geq 1, i \in S\}$ such that $\lim_{k_1 \to \infty} \hat{\varphi}_n'(k_1, i) = 0$, $\hat{\varphi}_n(k_1, i_0) \equiv 1$ for all $i \in S$, which, together with the boundedness of $|\hat{\varphi}_n(k_1, i)|$ in k_1 and the denumerability of $i \in S$, gives the existence of a subsequence $\{k_2\}$ of $\{k_1\}$ such that the limits

$$\lim_{k_2 \to \infty} \hat{\varphi}_n(k_2, i) =: \psi_n(i) \in [0, V_0(i)] \tag{26}$$

exist for all $i \in S$, and $\psi_n(i_0) = 1$.
Furthermore, Lemma 3.2(b) together with (26) guarantee the existence of a subsequence $\{k_3\}$ of $\{k_2\}$ ensuring the existence of the limit

$$\lim_{k_3 \to \infty} \frac{\varphi_n'(k_3, i_0)}{\varphi_n(k_3, i_0)} =: \rho_n \in [0, L(i_0)]. \tag{27}$$

Hence, along the sequence $\{k_3\}$ ($\subset \{k_2\} \subset \{k_1\} \subset \{k_0(i, m), m \geq 1, i \in S\}$), we have

$$\lim_{k_3 \to \infty} \left[\hat{\varphi}_n'(k_3, i) + \hat{\varphi}_n(k_3, i) \frac{\varphi_n'(k_3, i_0)}{\varphi_n(k_3, i_0)} \right] = \rho_n \psi_n(i) \qquad \text{for all } i \in S. \tag{28}$$

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

On the other hand, for any number k_3 in the sequence $\{k_3\}$, using the definition of $\hat{\varphi}_n(t, i)$ in (24) and Theorem 3.1(a) with (c, φ) replaced with the corresponding (c_n, φ_n), direct calculation gives

$$\hat{\varphi}_n{}'(k_3, i) + \hat{\varphi}_n(k_3, i)\frac{\varphi_n{}'(k_3, i_0)}{\varphi_n(k_3, i_0)} = \inf_{a\in A(i)}\left\{c_n(i, a)\hat{\varphi}_n(k_3, i) + \sum_{j\in S}\hat{\varphi}_n(k_3, j)q(j \mid i, a)\right\}$$
$$\le c_n(i, a)\hat{\varphi}_n(k_3, i) + \sum_{j\in S}\hat{\varphi}_n(k_3, j)q(j \mid i, a) \qquad \text{for all } (i, a) \in K, \tag{29}$$

which, together with the dominated convergence theorem and (26)–(28), imply

$$\rho_n\psi_n(i) \le \inf_{a\in A(i)}\left\{c_n(i, a)\psi_n(i) + \sum_{j\in S}\psi_n(j)q(j \mid i, a)\right\} \qquad \text{for all } i \in S.$$

Moreover, for each given k_3 and $i \in S$, by Assumption 3.3 and (28), there exists $a(i, k_3) \in A(i)$ (depending on k_3, $i \in S$) such that

$$\hat{\varphi}_n{}'(k_3, i) + \hat{\varphi}_n(k_3, i)\frac{\varphi_n{}'(k_3, i_0)}{\varphi_n(k_3, i_0)}$$
$$= c_n(i, a(i, k_3))\hat{\varphi}_n(k_3, i) + \sum_{j\in S}\hat{\varphi}_n(k_3, j)q(j \mid i, a(i, k_3)). \tag{30}$$

Since $A(i)$ is compact, there exist a subsequence $\{k_4\}$ of $\{k_3\}$ and $a'(i) \in A(i)$ such that $\lim_{k_4\to\infty} a(i, k_4) = a'(i)$. Thus, replacing k_3 in (30) with k_4 and then letting $k_4 \to \infty$, by (28) and (30) we have

$$\rho_n\psi_n(i) = c_n(i, a'(i))\psi_n(i) + \sum_{j\in S}\psi_n(j)q(j \mid i, a'(i))$$
$$\ge \inf_{a\in A(i)}\left\{c_n(i, a)\psi_n(i) + \sum_{j\in S}\psi_n(j)q(j \mid i, a)\right\},$$

which, together with (29), (26), and (27), gives (a).

(b) Fix any $i \ge n + 1$ with $n \ge i_0$, and let $\tau(n) := \tau_{\{0,1,\dots,i_0,\dots,n\}}$ be the hitting time. Then, by Lemma 3.1(b) and $\tau(n) \le \tau_{i_0}$, $\mathbb{P}_i^\pi(\tau(n) < \infty) = 1$. Moreover, since $\varphi_n(0, i) \equiv 1$ and $\varphi_n(t, i)$ is increasing in $t \ge 0$, using (4) and Theorem 3.1(c) (with c replacing by c_n here) as well as the fact that $c_n(x_v, \pi_v) \equiv 0$ for $v < \tau(n)$, we have, for each $\pi \in \Pi$,

$$\varphi_n(t, i) \le \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{t\wedge\tau(n)} c_n(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi_n(t - t\wedge\tau(n), x_{t\wedge\tau(n)})\right]$$
$$= \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{t\wedge\tau(n)} c_n(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi_n(t - t\wedge\tau(n), x_{t\wedge\tau(n)})\mathbf{1}_{\{t\le\tau(n)\}}\right]$$
$$+ \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{t\wedge\tau(n)} c_n(x_v, \pi_v)\,\mathrm{d}v\right\}\varphi_n(t - t\wedge\tau(n), x_{t\wedge\tau(n)})\mathbf{1}_{\{t>\tau(n)\}}\right]$$

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

$$
\begin{aligned}
&= \mathbb{E}_i^\pi\left[\varphi_n(0, x_t)\mathbf{1}_{\{t\le\tau(n)\}}\right] + \mathbb{E}_i^\pi\left[\varphi_n(t-\tau(n), x_{\tau(n)})\mathbf{1}_{\{t>\tau(n)\}}\right]\\
&\le 1 + \mathbb{E}_i^\pi\left[\varphi_n(t, x_{\tau(n)})\mathbf{1}_{\{t>\tau(n)\}}\right]\\
&\le 1 + \max\{\varphi_n(t, k), k = 0, 1, \dots, n\},
\end{aligned}
$$

which, together with Lemma 3.2(a) and $\varphi_n \ge 1$, imply that $\hat{\varphi}_n(t, i) \le 1 + \max\{V_0(k), k = 0, 1, \dots, n\} =: K(n)$ for all $t \ge 0$, $i \in S$. Hence, using (26) and Lemma 3.2(a) again, we have

$$
\psi_n(i) \le K(n) \qquad \text{for all } t \ge 0,\ i \in S. \tag{31}
$$

For each $\pi \in \Pi$ and $i \in S$, using Lemma A.1(a) and (5), we have

$$
\mathbb{E}_i^\pi\left[\exp\left\{\int_0^T\int_A [c_n(x_t, a) - \rho_n]\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right\}\psi_n(x_T)\right] - \psi_n(i) \ge 0 \qquad \text{for } T > 0.
$$

Hence, by (31) we have

$$
K(n)\mathrm{e}^{-T\rho_n}\mathbb{E}_i^\pi\left[\exp\left\{\int_0^T\int_A c_n(x_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right\}\right] \ge \psi_n(i).
$$

Taking logarithms of both sides, dividing by T, and letting $T \to \infty$, we obtain

$$
\begin{aligned}
\rho_n &\le \limsup_{T\to\infty}\frac{1}{T}\ln\mathbb{E}_i^\pi\left[\exp\left\{\int_0^T\int_A c_n(x_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right\}\right]\\
&\le \limsup_{T\to\infty}\frac{1}{T}\ln\mathbb{E}_i^\pi\left[\exp\left\{\int_0^T\int_A c(x_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right\}\right],
\end{aligned}
$$

which, together with the arbitrariness of π, completes the proof of (b).

(c) For each $\pi \in \Pi_{\mathrm{m}}^{\mathrm{d}}$ and $i \in S$, using Lemma A.1(b) and (5) again, we have

$$
\mathbb{E}_i^\pi\left[\exp\left\{\int_0^{T\wedge\tau_{i_0}}(c_n(x_t, \pi_t) - \rho_n)\,\mathrm{d}t\right\}\psi_n(x_{T\wedge\tau_{i_0}})\right] - \psi_n(i) \ge 0 \qquad \text{for } T > 0.
$$

Since ψ_n is bounded and $c_n(i, a) \le c(i, a) \le \delta(i)$ for all $(i, a) \in K$, by the dominated convergence theorem and Lemma 3.1(a), letting $T \to \infty$, we obtain

$$
\begin{aligned}
\psi_n(i) &\le \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{\tau_{i_0}}[c_n(x_t, \pi_t) - \rho_n]\,\mathrm{d}t\right\}\psi_n(x_{\tau_{i_0}})\right]\\
&= \mathbb{E}_i^\pi\left[\exp\left\{\int_0^{\tau_{i_0}}[c_n(x_t, \pi_t) - \rho_n]\,\mathrm{d}t\right\}\right] \le V_0(i),
\end{aligned} \tag{32}
$$

which, together with the arbitrariness of $\pi \in \Pi_{\mathrm{m}}^{\mathrm{d}}$, imply

$$
\psi_n(i) \le \inf_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{d}}}\mathbb{E}_i^\pi\left[\exp\left\{\int_0^{\tau_{i_0}}[c_n(x_t, \pi_t) - \rho_n]\,\mathrm{d}t\right\}\psi_n(x_{\tau_{i_0}})\right] \le V_0(i). \tag{33}
$$

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

On the other hand, take $f_n^* \in F$ to be a minimizing stationary policy in (5), that is,

$$\rho_n \psi_n(i) = c(i, f_n^*(i))\psi_n(i) + \sum_{j \in S} \psi_n(j) q(j \mid i, f_n^*(i)) \qquad \text{for all } i \in S. \tag{34}$$

Using Lemma A.1(b) and (34), as in the proof of (32) (by Fatou's lemma again) we have

$$\psi_n(i) \geq \mathbb{E}_i^{f_n^*} \left[\exp \left\{ \int_0^{\tau_{i_0}} (c_n(x_t, f_n^*(x_t)) - \rho_n) \, \mathrm{d}t \right\} \right].$$

This inequality, as well as (33), gives (c).

Remark A.1. In order to establish the existence of $k_0(m, i)$ for (25) to hold, the existence of $\varphi'(t, i)$ for the function $\varphi(t, i)$ at each $t \geq 0$ is required. Otherwise, such $k_0(m, i)$ cannot be guaranteed. For example, for some given $i \in S$ and $n \geq 1$, suppose that

$$\hat{\varphi}_n(t, i) := \begin{cases} t + 1, & t \in [1, 1.5), \\ 4 - t, & t \in [1.5, 2]. \end{cases}$$

Take $m = 1$. Then, $\frac{\hat{\varphi}_n(2,i) - \hat{\varphi}_n(1,i)}{2-1} = 0 \neq \hat{\varphi}_n{}'(k_0(1, i), i)$ for any $k_0(1, i) \in [1, 2]$.

Acknowledgements

Research supported by the National Natural Science Foundation of China (grant nos. 11931018, 61773411, 11961005) and Guangdong Province Key Laboratory of Computational Science at the Sun Yat-Sen University (2020B1212060032).

References

[1] Anderson, W. J. (1991). *Continuous-Time Markov Chains*. Springer, New York.
[2] Bäuerle, N. and Rieder, U. (2014). More risk-sensitive Markov decision processes. *Math. Operat. Res.* **39**, 105–120.
[3] Cavazos-Cadena, R. and Hernández-Hernández, D. (2011). Discounted approximations for risk-sensitive average criteria in Markov decision chains with finite state space. *Math. Operat. Res.* **36**, 133–146.
[4] Cavazos-Cadena, R. and Montes-de-Oca, R. (2000). Nearly optimal policies in risk-sensitive positive dynamic programming on discrete spaces. *Math. Meth. Operat. Res.* **52**, 133–167.
[5] Di Masi, G. B. and Stettner, L. (2007). Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM J. Control Optim.* **46**, 231–252.
[6] Feinberg, E. A., Mandava, M. and Shiryaev, A. N. (2014). On solutions of Kolmogorov's equations for nonhomogeneous jump Markov processes. *J. Math. Anal. Appl.* **411**, 261–270.
[7] Ghosh, M. K. and Saha, S. (2014). Risk-sensitive control of continuous time Markov chains. *Stochastics* **86**, 655–675.
[8] Guo, X. and Zhang, Y. (2020). On risk-sensitive piecewise deterministic Markov decision processes. *Appl. Math. Optim.* **81**, 685–710.
[9] Guo, X., Liu, Q. L. and Zhang, Y. (2019). Finite horizon risk-sensitive continuous-time Markov decision processes with unbounded transition and cost rates. *4OR* **17**, 427–442.
[10] Guo, X. P. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes*. Springer, Berlin.
[11] Guo, X. P. and Liao, Z. W. (2019). Risk-sensitive discounted continuous-time Markov decision processes with unbounded rates. *SIAM J. Control Optim.* **57**, 3857–3883.
[12] Guo, X. P. and Piunovskiy, A. (2011). Discounted continuous-time Markov decision processes with constraints: Unbounded transition and loss rates. *Math. Operat. Res.* **36**, 105–132.
[13] Guo, X. P. and Song, X. Y. (2011). Discounted continuous-time constrained Markov decision processes in Polish spaces. *Ann. Appl. Prob.* **21**, 2016–2049.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press

[14] GUO, X. P. AND ZHANG, J. Y. (2019). Risk-sensitive continuous-time Markov decision processes with unbounded rates and Borel spaces. *Discrete Event Dyn. Syst.* **29**, 445–471.
[15] GUO, X. P., HUANG, X. X. AND HUANG, Y. H. (2015). Finite-horizon optimality for continuous-time Markov decision processes with unbounded transition rates. *Adv. Appl. Prob.* **47**, 1064–1087.
[16] HOWARD, R. AND MATHESON, J. (1972). Risk-sensitive Markov decision proceses. *Manag. Sci.* **18**, 356–369.
[17] HUANG, Y. H., LIAN, Z. T. AND GUO, X. P. (2018). Risk-sensitive semi-Markov decision processes with general utilities and multiple criteria. *Adv. Appl. Prob.* **50**, 783–804.
[18] HUANG, Y. H., LIAN, Z. T. AND GUO, X. P. (2020). Risk-sensitive finite-horizon piecewise deterministic Markov decision processes. *Operat. Res. Lett.* **48**, 96–103.
[19] JAŚKIEWICZ, A. (2007). Average optimality for risk-sensitive control with general state space. *Ann. Appl. Prob.* **17**, 654–675.
[20] KAREL, S. (2018). Risk-sensitive average optimality in Markov decision processes. *Kybernetika (Prague)* **54**, 1218–1230.
[21] KITAEV, M. Y. AND RYKOV, V. V. (1995). *Controlled Queueing Systems*. CRC Press, New York.
[22] KUMAR, K. S. AND PAL, C. (2013). Risk-sensitive control of jump process on denumerable state space with near monotone cost. *Appl. Math. Optim.* **68**, 311–331.
[23] KUMAR, K. S. AND PAL, C. (2015). Risk-sensitive ergodic control of continuous-time Markov processes with denumerable state space. *Stochastic Anal. Appl.* **33**, 863–881.
[24] WEI, Q. D. (2016). Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion. *Math. Meth. Operat. Res.* **84**, 461–487.
[25] WEI, Q. D. AND CHEN, X. (2016). Continuous-time Markov decision processes under the risk-sensitive average cost criterion. *Operat. Res. Lett.* **44**, 457–462.
[26] WEI, Q. D. AND CHEN, X. (2017). Average cost criterion induced by the regular utility function for continuous-time Markov decision processes. *Discrete Event Dyn. Syst.* **27**, 501–524.
[27] WEI, Q. D. AND CHEN, X. (2019). Risk-sensitive average continuous-time Markov decision processes with unbounded rates. *Optimization* **68**, 773–800.
[28] ZHANG, Y. (2017). Continuous-time Markov decision processes with exponential utility. *SIAM J. Control Optim.* **55**, 2636–2660.

https://doi.org/10.1017/jpr.2020.105 Published online by Cambridge University Press