J Optim Theory Appl (2016) 170:670–686
DOI 10.1007/s10957-016-0916-z

Controlled Semi-Markov Chains with Risk-Sensitive Average Cost Criterion

Selene Chávez-Rodríguez[1] · **Rolando Cavazos-Cadena**[2] · **Hugo Cruz-Suárez**[1]

Received: 8 November 2015 / Accepted: 27 February 2016 / Published online: 11 March 2016

Abstract This work concerns with semi-Markov decision chains on a finite state space. Assuming that the controller has a constant and positive risk-sensitive coefficient, an optimality equation for the corresponding (long-run) risk-sensitive average cost index is formulated and, under suitable continuity-compactness conditions, it is shown that a solution of such an equation determines the optimal average cost, as well as an optimal stationary policy. Additionally, if the underlying Markov chain is communicating, then it is proved that the optimality equation has a solution.

Keywords Exponential utility · Certainty equivalent · Light tails · Communicating underlying chain · Risk-sensitive control

Mathematics Subject Classification 90C40 · 93E20 · 60J05

1 Introduction

This paper is concerned with semi-Markov decision chains evolving on a finite state space, which are mathematical models for a dynamical system whose state changes

✉ Rolando Cavazos-Cadena
rcavazos@uaaan.mx

Selene Chávez-Rodríguez
selenechavez@alumnos.fcfm.buap.mx

Hugo Cruz-Suárez
hcs@fcfm.buap.mx

1 Facultad de Ciencias Físico Matemáticas, Ave. San Claudio y Río Verde, Col. San Manuel CU, Benemérita Universidad Autónoma de Puebla, 72570 Puebla, PUE, Mexico

2 Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, 25315 Saltillo, COAH, Mexico

at random times. When the system arrives at one of the possible states, the controller applies an admissible action and such an intervention has several effects: The distribution of the random sojourn time at the state is determined, an immediate cost is incurred, a holding cost is continuously paid at a certain rate until a transition occurs, and the distribution of the next state depends on both the original one and the selected action. As the system evolves, the costs are accumulated and, assuming that the controller has a constant and positive risk-sensitivity coefficient, the performance of the control policy is measured by the corresponding (long-run) expected average cost criterion. The main objective of the work is *to set an optimality equation* characterizing the optimal risk-sensitive average cost and rendering an optimal stationary policy. This problem will be analyzed assuming that the random sojourn times are bounded, as well as standard continuity-compactness conditions formulated in Assumption 2.1 below. The main results of the paper can be roughly described as follows:

(i) *Verification result* If the optimality equation admits a solution, then the optimal risk-sensitive average cost function is constant and is immediately determined; moreover, as usual, an optimal stationary policy can be obtained by taking minimizers of the term within brackets in the right-hand side of the optimality equality.
(ii) *Existence of solutions* The optimality equation has a solution if, in addition to Assumption 2.1, the following condition is satisfied: Under any stationary policy, every state can be visited with positive probability regardless of the initial state.

Controlled semi-Markov chains have been widely used in applications, for instance, in the study of queueing systems [1–3], production scheduling [4], or in maintenance problems [5]. The average index has been intensively studied under the assumption that the controller is *risk-neutral*; this means that a random cost Y is assessed via its expected value. In that context, assuming that under the action of any stationary policy the underlying Markov chain has a single recurrent class, mild conditions ensure that the optimal risk-neutral average cost is constant and is characterized by a *single optimality equation*; see, for instance, [3,6,7]. On the other hand, the study of the risk-sensitive average criterion started, at least, with the seminal paper by Howard and Matheson [8]. Recent work on that index has been highly concentrated on *discrete-time* controlled Markov chains; models with finite state space have been studied; for instance, in [9–11], the case of a denumerable state space was examined in [12,13], whereas systems with general Borel state space were analyzed in [14–17]. Applications to mathematical finance are presented in [18], and risk-sensitive criteria with respect to general utilities are studied in [19].

Presently, the risk-sensitive average criterion for discrete-time Markov decision chains on a finite state space is well understood, and necessary and sufficient conditions for the characterization of the optimal average cost via a single optimality equation are known [11,20]. However, to the best of the authors' knowledge, a characterization of the optimal *risk-sensitive* average index via an optimality equation is not presently available in the semi-Markov context, a fact that provides the motivation for this work. The results in this paper extend conclusions recently obtained in [21], where *uncontrolled* semi-Markov chains were studied.

The organization of the paper is as follows: In Sect. 2, the controlled semi-Markov model is introduced, and the risk-sensitive average index is defined. Then, in Sect. 3,

the risk-sensitive average optimality equation is formulated, and the verification and existence results are stated as Theorems 3.1 and 3.2, respectively; the approach used to prove those conclusions, and the role of the basic assumptions in the arguments, are discussed in Remark 3.1. Next, Sect. 4 contains the auxiliary results that will be used to prove Theorem 3.1 in Sect. 5, whereas the demonstration of the existence theorem is presented in Sect. 6. The exposition concludes in Sect. 7 with a brief discussion about the results of the paper and an open problem.

2 The Model

In this section, the semi-Markov decision chain and the average criterion studied in the paper are introduced. First, it is convenient to state some basic notation.

Notation For a topological space W, the corresponding Borel σ-field is denoted by $\mathscr{B}(W)$, whereas $\mathbb{B}(W)$ stands for the class of all bounded and continuous functions defined on W. The supremum norm on $\mathbb{B}(W)$ is given by $\|h\| := \sup_{w\in W} |h(w)| < \infty$, for every $h \in \mathbb{B}(W)$. The indicator function corresponding to an event A is denoted by $I[A]$, and every relation involving random variables holds almost surely with respect to the underlying probability measure. Finally, for a sequence $\{r_k\}$ of real numbers, $\sum_{k=a}^{b} r_k := 0$ when $b < a$.

Throughout the remainder $\mathscr{S} := \big(S, A, \{A(x)\}, C, \{\rho_{x,a}(\cdot)\}, \{F_{x,a}\}, [p_{x,y}(\cdot)]\big)$ stands for a controlled semi-Markov chain. The components of $\mathscr{S}$ are as follows: The *finite* set S is the *state space* and is endowed with the discrete topology, the metric space A is the *action set* and, for every $x \in S$, $A(x) \subset A$ is the (nonempty) subset of admissible actions at x. On the other hand, $C : \mathbb{K} \to \mathbb{R}$ is the immediate cost function, where $\mathbb{K} := \{(x, a) : a \in A(x), x \in S\}$ is the class of admissible pairs, whereas for each $(x, a) \in \mathbb{K}$ the mappings $\rho_{x,a} : [0, \infty[\to \mathbb{R}$ and $F_{x,a}(\cdot)$ are the holding cost rate and the sojourn time distribution function, respectively, corresponding to the application of action $a \in A(x)$ at state x; it is assumed that the sojourn times are positive, so that

$$F_{x,a}(0) = 0, \quad (x, a) \in \mathbb{K}. \tag{1}$$

Finally, $[p_{x,y}(a)]$ is the (controlled) transition law and satisfies $\sum_{y\in S} p_{x,y}(a) = 1$ for every $(x, a) \in \mathbb{K}$. The model $\mathscr{S}$ has the following interpretation: At time $t = 0$ the system starts at $X_0 = x_0 \in S$. Now, suppose that after completing the nth transition the system arrives at state $X_n = x$. At the arrival time, the decision maker applies a control (action) $A_n = a \in A(x)$ and such an intervention has four consequences: (i) A cost $C(x, a)$ is incurred, (ii) the system stays at x during a (random) sojourn time S_n whose distribution function is $F_{x,a}$, (iii) a holding cost is incurred at a rate $\rho_{x,a}$, while the system stays at x, and (iv) regardless of the sojourn time S_n and the previous states, actions and sojourn times, after S_n has elapsed the system jumps to other state $X_{n+1} = y \in S$ with probability $p_{x,y}(a)$; this is the Markov property of the decision process. Note that the nth transition is completed at time T_n, where

$$T_0 = 0 \quad \text{and} \quad T_n = \sum_{i=0}^{n-1} S_i, \quad n = 1, 2, 3, \ldots, \tag{2}$$

so that the number of transitions N_t in the interval $[0, t]$ is given by

$$N_t = \sup\{n \in \mathbb{N} : T_n \leq t\}, \quad t \geq 0. \tag{3}$$

The information recorded by the controller up to time T_n is given by $\mathscr{H}_n$, where

$$\mathscr{H}_0 := X_0, \text{ and } \mathscr{H}_n := (X_0, A_0, S_0, \ldots, X_{n-1}, A_{n-1}, S_{n-1}, X_n), \quad n \geq 1, \tag{4}$$

so that, for every $t \geq 0$ and $n \in \mathbb{N}$,

$$T_n \text{ is } \sigma(\mathscr{H}_n)\text{-measurable, and } [N_t \geq n] = [T_n \leq t] \in \sigma(\mathscr{H}_n). \tag{5}$$

Assumption 2.1 (i) For each $x \in S$, the set $A(x)$ is a compact subspace of A.
(ii) For each $x, y \in S$, $a \mapsto C(x, a)$ and $a \mapsto p_{xy}(a)$ are continuous in $a \in A(x)$.
(iii) The family $\{F_{x,a}\}_{(x,a)\in\mathbb{K}}$ is supported on a compact interval and is weakly continuous, that is, there exists $B > 0$ such that

$$F_{x,a}(B) = 1, \quad (x, a) \in \mathbb{K}, \tag{6}$$

and for each $x \in S$ and $u \in \mathbb{B}([0, B])$, $a \mapsto \int_0^B u(s)\, \mathrm{d}F_{x,a}(S)$ is continuous in $a \in A(x)$.
(iv) For every $x \in S$, the mapping $(a, s) \mapsto \rho_{x,a}(s)$ is continuous in $(a, s) \in A(x) \times [0, B]$.

Except for the requirement (6), this assumption is rather standard. The role of condition (6) will be discussed in Remark 3.1(ii). Since the space $A(x)$ is compact so is $A(x) \times [0, B]$; thus, $\sup_{(a,s)\in A(x)\times[0,B]} |\rho_{x,a}(s)| < \infty$ for every $x \in S$, by part (iv) of Assumption 2.1, and then, using that the state space is finite, it follows that

$$B_\rho := \sup_{(x,a)\in\mathbb{K},\ s\in[0,B]} |\rho_{x,a}(s)| < \infty. \tag{7}$$

Policies A policy is a rule for choosing actions which, at each decision epoch T_n, may depend on the current state as well as on the record of previous states, actions and sojourn times. A more formal description is as follows: For each $n = 0, 1, 2, \ldots$, define the space $\mathbb{H}_n$ of admissible histories until the completion of nth transition by $\mathbb{H}_0 := S$, and $\mathbb{H}_n := \mathbb{K}\times]0, \infty[\times\mathbb{H}_{n-1}$ for $n = 1, 2, 3, \ldots$. A generic element of $\mathbb{H}_n$ is denoted by $h_n = (x_0, a_0, s_0, x_1, \ldots, x_{n-1}, a_{n-1}, s_{n-1}, x_n)$, where $x_i \in S$, $a_i \in A(x_i)$ and $s_i > 0$. A *control policy* $\pi = \{\pi_n\}$ is a special sequence of stochastic kernels: For each $n \in \mathbb{N}$ and $h_n \in \mathbb{H}_n$, $\pi_n(\cdot \,|\, h_n)$ is a probability measure on $\mathscr{B}(A)$ satisfying that $\pi_n(A(x_n) \,|\, h_n) = 1$, whereas for every $B \in \mathscr{B}(A)$, the mapping $h_n \mapsto \pi_n(B \,|\, h_n)$ is Borel measurable. After observing the event $[\mathscr{H}_n = h_n]$, under the action of policy π the probability of choosing the nth action A_n within $B \in \mathscr{B}(A)$ is given by $\pi_n(B \,|\, h_n)$.

The collection of all policies is denoted by $\mathscr{P}$. Define $\mathbb{F} := \prod_{x\in S} A(x)$, which is a compact metric space and consists of all functions $f : S \to A$ such that $f(x) \in A(x)$ for each $x \in S$. A policy $\pi \in \mathscr{P}$ is *stationary* if there exists $f \in \mathbb{F}$ such that the equality $\pi_n(\{f(x_n)\} \,|\, h_n) = 1$ is always valid; in this case π and f are naturally identified and, with this convention, $\mathbb{F} \subset \mathscr{P}$. Given the initial state $X_0 = x$ and the policy π being used for choosing actions, the distribution of $\{(X_n, A_n, S_n)\}_{n\in\mathbb{N}}$ is uniquely determined via the Tulcea theorem [22]. Such a distribution is denoted by P_x^π, whereas E_x^π stands for the corresponding expectation operator. The following Markov relations are satisfied almost surely under each distribution P_x^π: For each $x, y \in S$, $\tilde{A} \in \mathscr{B}(A)$ and $n \in \mathbb{N}$,

$$\begin{aligned} P_x^\pi[X_0 = x] &= 1, \\ P_x^\pi[A_n \in \tilde{A} \,|\, \mathscr{H}_n] &= \pi_n(\tilde{A} \,|\, \mathscr{H}_n), \\ P_x^\pi[S_n \le t \,|\, \mathscr{H}_n, A_n] &= F_{X_n,A_n}(t), \quad t \ge 0, \\ P_x^\pi[X_{n+1} = y \,|\, \mathscr{H}_n, A_n, S_n] &= p_{X_n,y}(A_n). \end{aligned} \tag{8}$$

As it will be shown in Lemma 4.1 below, the third equality in this display and (1), together yield that N_t is finite P_x^π almost surely for each $t > 0$, $x \in S$ and $\pi \in \mathscr{P}$.

The total cost up to a positive time Suppose that the system will be driven by the controller up to time $t > 0$, so that the states $X_0, X_1, \ldots, X_{N_t}$ will be visited at times $T_0, T_1, \ldots, T_{N_t}$, respectively. For each nonnegative integer $k \le N_t$, the action A_k will be applied at X_k, incurring an immediate cost $C(X_k, A_k)$. As for the holding costs, note that for $k < N_t$, the system will stay at X_k during the interval $[T_k, T_{k+1}[\subset [0, t]$, where the inclusion is due to the fact that $0 \le T_{k+1} \le T_{N_t} \le t$, by (2) and (3). Since $T_{k+1} = T_k + S_k$, it follows that the system stays at X_k during S_k units of time, incurring the holding cost $\int_0^{S_k} \rho_{X_k,A_k}(r)\,\mathrm{d}r$. On the other hand, at time T_{N_t} the system arrives at state X_{N_t} and stays there during the interval $[T_{N_t}, t]$, since the next transition will occur at time $T_{N_t+1} > t$; thus within the observation interval $[0, t]$, the system stays at X_{N_t} in an interval of length $t - T_{N_t}$, with corresponding holding cost $\int_0^{t-T_{N_t}} \rho_{X_{N_t},A_{N_t}}(r)\,\mathrm{d}r$. Therefore, the total cost incurred up to time $t > 0$ is given by

$$\begin{aligned} \mathscr{C}_t := \sum_{k=0}^{N_t-1} \left[C(X_k, A_k) + \int_0^{S_k} \rho_{X_k,A_k}(r)\,\mathrm{d}r \right] \\ + C(X_{N_t}, A_{N_t}) + \int_0^{t-T_{N_t}} \rho_{X_{N_t},A_{N_t}}(r)\,\mathrm{d}r. \end{aligned} \tag{9}$$

Since N_t is always finite with probability 1, so is $\mathscr{C}_t$, by (6) and (7).

Utility function and average criterion Throughout the remainder it is supposed that the controller has constant risk-sensitivity $\lambda > 0$; this means that a random cost Y is assessed via $E[U_\lambda(Y)]$ (assumed to be finite), where $U_\lambda(\cdot)$ is the exponential utility given by $U_\lambda(x) := e^{\lambda x}$ for $x \in \mathbb{R}$. If the decision maker can choose between incurring one of the random costs Y_1 or Y_0, then paying Y_0 will be preferred when $E[U_\lambda(Y_1)] > E[U_\lambda(Y_0)]$, whereas the controller will be indifferent between Y_1 and

Y_0 when $E[U_\lambda(Y_1)] = E[U_\lambda(Y_0)]$. The *certainty equivalent* of a random cost Y with respect to U_λ is the real number $\mathscr{E}_\lambda[Y]$ given by

$$\mathscr{E}_\lambda[Y] := \frac{1}{\lambda} \log \left(E \left[e^{\lambda Y} \right] \right); \tag{10}$$

since $U_\lambda(\mathscr{E}_\lambda[Y]) = E[U_\lambda(Y)]$, the controller is willing to pay the nonrandom amount $\mathscr{E}_\lambda[Y]$ to avoid facing the uncertainty conveyed by Y. Now, suppose that the system will be driven up to time $t > 0$ using policy $\pi \in \mathscr{P}$ starting at $X_0 = x$. The total cost incurred in the time interval $[0, t]$ is given in (9), and instead of facing the random amount $\mathscr{C}_t$, the decision maker will gladly pay the certainty equivalent

$$J_{t,\lambda}(x, \pi) := \frac{1}{\lambda} \log \left(E_x^\pi \left[e^{\lambda \mathscr{C}_t} \right] \right), \tag{11}$$

which represents an average of $J_{t,\lambda}(x, \pi)/t$ per unit of time. The λ-sensitive average cost at state x under π is the largest limit point of those averages as t goes to ∞ :

$$J_\lambda(x, \pi) := \limsup_{t \to \infty} \frac{1}{t} J_{t,\lambda}(x, \pi). \tag{12}$$

The optimal (λ-sensitive) average cost at state x is

$$J_\lambda^*(x) := \inf_{\pi \in \mathscr{P}} J_\lambda(x, \pi), \tag{13}$$

and a policy $\pi^* \in \mathscr{P}$ is (λ-)average optimal if $J_\lambda(x, \pi) = J_\lambda^*(x)$ for every state x.

The problem The main objective of the paper can be now stated as follows:

- To determine an *optimality equation* whose solutions characterize the optimal average cost function $J_\lambda^*(\cdot)$ and render optimal policies.

This problem is twofolded: On the one hand, it must be proved that J_λ^* can be obtained from a solution of the optimality equation (the verification result) and, on the other hand, conditions must be provided ensuring the existence of solutions. In the following sections the optimality equation for the risk-sensitive average index will be stated, and the verification result will be established under Assumption 2.1, whereas the existence theorem will be derived under an additional condition on the underlying discrete-time process $\{X_n\}$, namely, the communication property in Assumption 3.1 introduced in the following section.

3 Main Results

In this section the main conclusions of the paper are stated. Consider the equation

$$e^{\lambda h(x)} = \inf_{a \in A(x)} E_x \left[e^{\lambda \left[C(X_0, A_0) + \int_0^{S_0} \rho_{X_0, A_0}(t)\, \mathrm{d}t - g S_0 + h(X_1) \right]} \middle| A_0 = a \right], \quad x \in S, \tag{14}$$

where g is a real number and $h(\cdot)$ is a real function defined on the state space S. In terms of the certainty equivalents introduced in (10), this relation can be expressed as

$$h(x) = \mathscr{E}_\lambda\left[C(X_0, A_0) + \int_0^{S_0} \rho_{X_0,A_0}(t)\,\mathrm{d}t - gS_0 + h(X_1)|X_0 = x, A_0 = a\right], \quad x \in S, \tag{15}$$

whereas, via (8), the equality (14) can be more explicitly written as

$$e^{\lambda h(x)} = \inf_{a\in A(x)}\left[e^{\lambda C(x,a)}\int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\,\mathrm{d}t - gs]}\,\mathrm{d}F_{x,a}(s) \times \sum_{y\in S} p_{x,y}(a)e^{\lambda h(y)}\right], \quad x \in S. \tag{16}$$

As it is shown in the following theorem, each of these equivalent equalities is an *optimality equation* for the λ-sensitive average criterion.

Theorem 3.1 (Verification) *Suppose that the equality (16)* is satisfied by the pair $(g, h(\cdot))$. Under Assumption 2.1 the following assertions (i) and (ii) hold.

(i) $J^*_\lambda(x) = g$ for every $x \in S$.
(ii) For each $x \in S$, the term within brackets in the right-hand side of (16) has a minimizer $f^*(x) \in A(x)$, and the stationary policy f^* is optimal, that is, $J_\lambda(\cdot, f^*) = g$. Moreover, $\lim_{t\to\infty} \frac{1}{t} J_{t,\lambda}(x, f^*) = g$ for every $x \in S$.

It is interesting to observe that the right-hand side of (16) engages the *whole* distribution function $F_{x,a}$ of the sojourn time S_0 given that action a is applied at the state x. In contrast, the optimality equation in the risk-neutral context involves only *the expectations* of the holding cost $\int_0^{S_0} \rho_{x,a}(t)\,\mathrm{d}t$ and S_0 [3,6]. As noted in [21], a reason behind this difference is the nonlinearity of the mapping $Y \mapsto \mathscr{E}_\lambda[Y]$.

The existence of a solution of the optimality (16) will be established when, in addition to Assumption 2.1, the following condition is satisfied.

Assumption 3.1 Under each stationary policy, the Markov chain $\{X_n\}$ is communicating , that is, given $f \in \mathbb{F}$ and $x, y \in S$, there exists a positive integer $n_{x,y,f} \equiv n$ and states x_k, $0 \le k \le n$, such that (a) $x_0 = x$, $x_n = y$, and (b) $p_{x_{k-1},x_k}(f(x_{k-1})) > 0$ for $1 \le k \le n$,

Theorem 3.2 (Existence of solutions) *Under Assumptions 2.1 and 3.1, there exist* $g \in \mathbb{R}$ *and* $h : S \to \mathbb{R}$ *such that the optimality Eq. (16)* is satisfied.

Remark 3.1 (i) The existence of solutions of (16) cannot be generally ensured under the sole Assumption 2.1. Indeed, when the sojourn times S_k are identically 1, it is known that, if Assumption 3.1 fails, then the optimal average cost function $J^*_\lambda(\cdot)$ is not necessarily constant, and then (16) does not have a solution [11,20].

(ii) Given a solution $(g, h(\cdot))$ of (16), the main idea behind the proof of Theorem 3.1 is to compare the certainty equivalents of the *relative* total cost $\mathscr{C}_t - tg$, incurred up to time $t > 0$, with the one corresponding to $\tilde{\mathscr{C}}_t - gT_{N_t+1}$, where $\tilde{\mathscr{C}}_t$ is the total cost incurred up to the completion of the first transition *posterior* to t, which occurs at time T_{N_t+1}. Condition (6) in Assumption 2.1 makes it possible to compare both certainty equivalents in a neat way, and allows a streamlined exposition. As in [21], at the expense of complicating the argument, (6) can be replaced by requirements on the conditional distribution of the sojourn times $S_n - r$ given that $S_n > r$ for $r > 0$.

(iii) The argument used to derive Theorem 3.1 relies heavily on the following consequence of the Markov property: For each $t > 0$ the random variable N_t has 'light tails'; that is, regardless of the initial state and the policy employed, the tails of the distribution of N_t decay faster than any geometric sequence. On the other hand, the existence result will be proved combining (i) known results on the existence of solutions of the risk-sensitive optimality equation in *the discrete-time case* and (ii) the intermediate value property of a continuous mapping defined on an interval of the real line.

4 Auxiliary Tools for the Verification Theorem

This section contains the technical preliminaries that will be used to establish Theorem 3.1. The main tool is Theorem 4.1, whose proof relies on the following lemma.

Lemma 4.1 *Under Assumption 2.1 the assertions (i)–(iii) below hold:*

(i) Let $x \in S$ be arbitrary and suppose that the function $R : A(x) \times [0, B] \to \mathbb{R}$ is continuous. In this case, the mapping $a \mapsto \int_0^B R(a, s)dF_{x,a}(s)$ is continuous in $a \in A(x)$.

(ii) Given $\alpha \in]0, 1[$, there exists an integer $m_\alpha > 0$ such that, for every $(x, a) \in \mathbb{K}$, the inequality $\int_0^B e^{-\mu s}\, dF_{x,a}(s) \leq \alpha$ holds for every $\mu \geq m_\alpha$.

(iii) For each $\alpha \in]0, 1[$, $t \geq 0$ and $n \in \mathbb{N}$, $P_x^\pi [N_t \geq n] \leq \alpha^n e^{m_\alpha t}$ for all $x \in S$, and $\pi \in \mathscr{P}$, where m_α is as in part (ii), and then

$$P_x^\pi [N_t < \infty] = 1. \tag{17}$$

Proof (i) Given an arbitrary state x, let $\{a_n\} \in A(x)$ be a convergent sequence, write $a^* := \lim_{n\to\infty} a_n$, and note that $R(a^*, \cdot)$ is continuous in $[0, B]$, so that

$$\lim_{n\to\infty} \int_0^B R(a^*, s)\mathrm{d}F_{x,a_n}(s) = \int_0^B R(a^*, s)\mathrm{d}F_{x,a^*}(s),$$

by Assumption 2.1(iii). Observe now that $\|R(a_n, \cdot) - R(a^*, \cdot)\| \to 0$, by the 'tube lemma' in [23], and then

$$\left| \int_0^B R(a_n, s)\mathrm{d}F_{x,a_n}(s) - \int_0^B R(a^*, s)\mathrm{d}F_{x,a_n}(s) \right| \leq \|R(a_n, \cdot) - R(a^*, \cdot)\| \to 0.$$

Via the triangle inequality, these two last displays together imply that

$$\lim_{n\to\infty}\int_0^B R(a_n,s)\mathrm{d}F_{x,a_n}(s)=\int_0^B R(a^*,s)\mathrm{d}F_{x,a^*}(s),$$

and the desired conclusion follows.

(ii) Let $x\in S$ and $\alpha\in]0,1[$ be arbitrary but fixed. For each $n\in\mathbb{N}$ define the function $g_n:A(x)\to]0,\infty[$ by $g_n(a):=\int_0^B e^{-ns}\mathrm{d}F_{x,a}(s),\ a\in A(x)$, and note that $g_n(\cdot)$ is continuous, by Assumption 2.1(iii), whereas (1) and the dominated convergence theorem together yield that $\lim_{n\to\infty}g_n(a)=0$ for every $a\in A(x)$. Since $g_n(\cdot)\geq g_{n+1}(\cdot)$, the compactness of the action set $A(x)$ implies, that $\lim_{n\to\infty}\|g_n(\cdot)\|=0$, by Dini's theorem. Thus, there exists $m_{x,\alpha}>0$ such that $g_{m_{x,\alpha}}(a)=\int_0^B e^{-m_{x,\alpha}s}\mathrm{d}F_{x,a}(s)\leq\alpha$ for every $a\in A(x)$. Since S is finite, it follows that $m_\alpha:=\max\{m_{x,\alpha}:x\in S\}<\infty$, and the above display yields that $\int_0^B e^{-\mu s}\mathrm{d}F_{x,a}(s)\leq\alpha$ for every $(x,a)\in\mathbb{K}$ when $\mu\geq m_\alpha$.

(iii) Let $x\in S$, $\pi\in\mathscr{P}$ and $\alpha\in]0,1[$ be arbitrary. Given an integer $n>0$, the third equality in (8) yields that, conditionally on X_k, A_k for $k=0,1,2,\ldots,n-1$, the sojourn times $S_0,S_1,\ldots,S_{n-1}$ are independent with corresponding distribution functions $F_{X_0,A_0},F_{X_1,A_1},\ldots,F_{X_{n-1},A_{n-1}}$, respectively. Thus, with m_α as in part (ii), it follows that

$$E_x^\pi\left[e^{-m_\alpha[S_0+S_1+\cdots+S_{n-1}]}\,\middle|\,X_k,A_k,\ 0\leq k<n\right]=\prod_{k=0}^{n-1}\int_0^B e^{-m_\alpha s}\mathrm{d}F_{X_k,A_k}(s)\leq\alpha^n,$$

and then $E_x^\pi\left[e^{-m_\alpha T_n}\right]\leq\alpha^n$; see (2). This relation yields that, for any $t\geq 0$,

$$e^{-m_\alpha t}P_x^\pi[T_n\leq t]\leq E_x^\pi\left[e^{-m_\alpha T_n}\right]\leq\alpha^n.$$

Hence, the equality $[N_t\geq n]=[T_t\leq t]$ in (5) yields that $P_x^\pi[N_t\geq n]\leq\alpha^n e^{m_\alpha t}$ for every $n\in\mathbb{N}$; since $0<\alpha<1$, it follows that $P_x^\pi[N_t=\infty]=\lim_{n\to\infty}P_x^\pi[N_t\geq n]=0$. □

Theorem 4.1 *Let $(g,h(\cdot))$ be a solution of equation (16). Under Assumption 2.1, the assertions (i) and (ii) below hold:*

(i) The following inequality is valid for every $x\in S$, $\pi\in\mathscr{P}$ and $t>0$:

$$e^{\lambda h(x)}\leq E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(s)\,ds\right)-gT_{N_t+1}+h(X_{N_t+1})\right]}\right].\tag{18}$$

(ii) For each $x\in S$, the term within brackets in (16) has a minimizer $f^*(x)\in A(x)$, and the policy $f^*\in\mathbb{F}$ satisfies that, for every $x\in S$ and $t>0$,

$$e^{\lambda h(x)} = E_x^{f^*}\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(s)\,ds\right)-gT_{N_t+1}+h(X_{N_t+1})\right]}\right]. \quad (19)$$

Proof (i) Let $x \in S$ and $\pi \in \mathscr{P}$ be arbitrary but fixed, and note that (16) yields that for every $a \in A(x)$,

$$\begin{aligned} e^{\lambda h(x)} &\le e^{\lambda C(x,a)}\int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\,\mathrm{d}t-gs]}\,\mathrm{d}F_{x,a}(s)\sum_{y\in S}p_{x,y}(a)e^{\lambda h(y)} \\ &= E_x^{\pi}\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,\mathrm{d}t-gS_0+h(X_1)\right]}\,\middle|\, X_0=x,\, A_0=a\right]. \quad (20)\end{aligned}$$

More generally, via the Markov equalities (8) it follows that for every $n \in \mathbb{N}$,

$$e^{\lambda h(X_n)} \le E_x^{\pi}\left[e^{\lambda\left[C(X_n,A_n)+\int_0^{S_n}\rho_{X_n,A_n}(t)\,\mathrm{d}t-gS_n+h(X_{n+1})\right]}\,\middle|\,\mathscr{H}_n,\,A_n\right]; \quad (21)$$

see (4). It will be proved, by induction, that for every nonnegative integer n,

$$\begin{aligned} e^{\lambda h(x)} \le\ & E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{N_t+1}+h(X_{N_t+1})\right]}I[N_t\le n]\right] \\ &+E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}+h(X_{n+1})\right]}I[N_t>n]\right]. \quad (22)\end{aligned}$$

To achieve this goal, note that taking the integral with respect to $\pi_0(\cdot|x)$ in (20) it follows that

$$\begin{aligned} e^{\lambda h(x)} &\le E_x^{\pi}\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,\mathrm{d}t-gS_0+h(X_1)\right]}\right] \\ &= E_x^{\pi}\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,\mathrm{d}t-gS_0+h(X_1)\right]}I[N_t=0]\right] \\ &\quad +E_x^{\pi}\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,\mathrm{d}t-gS_0+h(X_1)\right]}I[N_t>0]\right]; \end{aligned}$$

since $T_1 = S_0$, this last display immediately yields that (22) is valid for $n = 0$. Now suppose that (22) holds for a certain nonnegative integer n, and observe that

$$\begin{aligned} &e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}+h(X_{n+1})\right]}I[N_t>n] \\ &= e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}\right]}I[N_t\ge n+1]e^{\lambda h(X_{n+1})} \end{aligned}$$

$$
\begin{aligned}
&\le e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}\right]}I[N_t\ge n+1]\\
&\quad\times E_x^{\pi}\left[e^{\lambda\left[C(X_{n+1},A_{n+1})+\int_0^{S_{n+1}}\rho_{X_{n+1},A_{n+1}}(t)\,\mathrm{d}t-gS_{n+1}+h(X_{n+2})\right]}\,\middle|\,\mathscr{H}_{n+1},A_{n+1}\right]\\
&\le E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-g[T_{n+1}+S_{n+1}]+h(X_{n+2})\right]}\right.\\
&\quad\left.\times I[N_t\ge n+1]\,|\,\mathscr{H}_{n+1},A_{n+1}\right],
\end{aligned}
$$

where (21) was used to set the first inequality, whereas the fact that the variables $I[N_t \ge n+1]$ and $\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})$ are $\sigma(\mathscr{H}_{n+1})$-measurable was used in the last step. Since $T_{n+2}=T_{n+1}+S_{n+1}$, by (2), it follows that

$$
\begin{aligned}
&E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}+h(X_{n+1})\right]}I[N_t>n]\right]\\
&\le E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t\ge n+1]\right]\\
&= E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t=n+1]\right]\\
&\quad+E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t>n+1]\right]\\
&= E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{N_t+1}+h(X_{N_t+1})\right]}I[N_t=n+1]\right]\\
&\quad+E_x^{\pi}\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t>n+1]\right],
\end{aligned}
$$

and combining this relation with the induction hypothesis it follows that (22) is also valid with $n+1$ instead of n, completing the induction argument. Next, observe that the monotone convergence theorem and (17) together yield that

$$\lim_{n\to\infty} E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} I[N_t \le n] \right]$$
$$= E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} \right]. \tag{23}$$

Now, a glance at (2) and (6) shows that the inequalities $S_k \le B$ and $T_{n+1} \le (n+1)B$ are always valid with probability 1, so that

$$e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n]$$
$$\le e^{\lambda(n+1)[\|C\|+B(B_\rho+|g|)]+\lambda\|h\|} I[N_t \ge n+1],$$

and then

$$E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n] \right]$$
$$\le e^{\lambda(n+1)[\|C\|+B(B_\rho+|g|)]+\lambda\|h\|} P_x^\pi[N_t \ge n+1].$$

Now set $\alpha = e^{-\lambda[\|C\|+B(B_\rho+|g|)]}/2$. Combining the above display and Lemma 4.1(ii) it follows that

$$E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,\mathrm{d}t\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n] \right]$$
$$\le e^{\lambda\|h\|}(1/2)^{n+1} e^{m_\alpha t} \to 0 \text{ as } n \to \infty.$$

After taking the limit as n goes to ∞ in both sides of (22), this last convergence and (23) together lead to (18).

(ii) Let $x \in S$ be arbitrary, and note that Assumption 2.1(iv) and the bounded convergence theorem together imply that $R(a,s) := e^{\lambda[\int_0^s \rho_{x,a}(t)\,\mathrm{d}t - gs]}$ is a continuous function of $(a,s) \in A(x)\times[0,B]$, and then the mapping $a \mapsto \int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\,\mathrm{d}t-gs]}\,dF_{x,a}(s)$ is continuous in its domain $A(x)$, by Lemma 4.1(i). Combining this property with Assumption 2.1(i), it follows that the term within brackets in (16) is a continuous function of $a \in A(x)$, and then the compactness of $A(x)$ ensures the existence of a minimizer $f^*(x) \in A(x)$. It follows that the stationary policy f^* satisfies

$$e^{\lambda h(x)} = e^{\lambda C(x,f^*(x))} \int_0^B e^{\lambda[\int_0^s \rho_{x,f^*(x)}(t)\,\mathrm{d}t - gs]}\,\mathrm{d}F_{x,f^*(x)}(s) \sum_{y\in S} p_{x,y}(f^*(x)) e^{\lambda h(y)}$$
$$= E_x^{f^*}\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,\mathrm{d}t - gS_0+h(X_1)\right]} \right], \quad x \in S.$$

Starting from this equality, (19) can be obtained paralleling the induction argument used in part (i) to obtain (18) from (20). □

5 Proof of the Verification Theorem

In this section Theorem 3.1 will be established. To begin with, note that (2) and (3) together yield that $T_{N_t} \leq t < T_{N_t+1} = T_{N_t} + S_{N_t}$, for every $t > 0$, and then

$$0 \leq t - T_{N_t} \leq S_{N_t} \leq B \text{ and } T_{N_t+1} - t \leq S_{N_t} \leq B; \tag{24}$$

see (6) for the right-most inequalities. Now, a glance at (9) yields that

$$\begin{aligned}\sum_{k=0}^{N_t}\left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k,A_k}(s)\,\mathrm{d}s\right) - gT_{N_t+1} \\ = (\mathscr{C}_t - tg) + \int_{t-T_{N_t}}^{S_{N_t}} \rho_{X_{N_t},A_{N_t}}(s)\,\mathrm{d}s - (T_{N_t+1} - t)g,\end{aligned}$$

and combining this equality with (7) and (24) it follows that

$$\left|\sum_{k=0}^{N_t}\left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k,A_k}(s)\,\mathrm{d}s\right) - gT_{N_t+1} - (\mathscr{C}_t - tg)\right| \leq B(B_\rho + |g|). \tag{25}$$

Proof of Theorem 3.1 Let $(x, \pi) \in S \times \mathscr{P}$ be arbitrary and note that (18) yields that $e^{-2\lambda\|h\|} \leq E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(s)\,\mathrm{d}s\right)-gT_{N_t+1}\right]}\right]$. Using (25), this leads to $e^{-2\lambda\|h\|} \leq E_x^\pi\left[e^{\lambda[\mathscr{C}_t - tg + B(B_\rho+|g|)]}\right]$, so that $e^{-2\lambda[\|h\| - B(B_\rho+|g|)+tg]} \leq E_x^\pi\left[e^{\lambda\mathscr{C}_t}\right]$, and then

$$g \leq \liminf_{t\to\infty} \frac{1}{\lambda t}\log\left(E_x^\pi\left[e^{\lambda\mathscr{C}_t}\right]\right) \leq J_\lambda(x, \pi); \tag{26}$$

since the policy π is arbitrary, it follows that

$$g \leq J_\lambda^*(x). \tag{27}$$

Next, let the policy $f^* \in \mathbb{F}$ be such that, for each $x \in S$, the action $f^*(x)$ is a minimizer of the right-hand side of (16). In this case, the equality (19) established in Theorem 4.1(ii) implies that $e^{2\lambda\|h\|} \geq E_x^{f^*}\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(s)\,ds\right)-gT_{N_t+1}\right]}\right]$; together with (25) this yields that $e^{2\lambda\|h\|} \geq E_x^{f^*}\left[e^{\lambda[\mathscr{C}_t - tg - B(B_\rho+|g|)]}\right]$, so that

$$e^{2\lambda\|h\|+\lambda B(B_\rho+|g|)+\lambda tg} \geq E_x^{f^*}\left[e^{\lambda\mathscr{C}_t}\right],$$

and then $g \geq \limsup_{t\to\infty} \frac{1}{\lambda t} \log\left(E_x^{f^*}\left[e^{\lambda \mathscr{C}_t}\right]\right) = J_\lambda(x, f^*)$. Combining this relation with (26) and (27), it follows that $J_\lambda^*(x) = g$ and $J_\lambda(x, f^*) = \lim_{t\to\infty} \frac{1}{\lambda t} E_x^{f^*}\left[e^{\lambda \mathscr{C}_t}\right]$; this completes the proof, since the state x is arbitrary. □

6 Existence of Solutions

In this section Theorem 3.2 will established. The argument is based on the risk-sensitive average criterion for the discrete-time process $\{X_n\}$. Let $\tilde{\mathscr{P}}$ be the subset of $\mathscr{P}$ that consists of all policies satisfying that, for each positive integer n, the kernel π_n depends only on the states $X_0, A_0, X_1, A_1, \ldots, X_{n-1}, A_{n-1}, X_n$, that is, for every history $h_n \in \mathbb{H}_n$, $\pi_n(\cdot|h_n) = \pi_n(\cdot|x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n)$. Given a function $D \in \mathbb{B}(\mathbb{K})$, define the discrete-time average index at $x \in S$ under $\pi \in \tilde{\mathscr{P}}$ by

$$V_{\lambda,D}(x,\pi) := \limsup_{n\to\infty} \frac{1}{\lambda n} \log\left(E_x^{\pi}\left[e^{\lambda \sum_{k=0}^{n-1} D(X_k, A_k)}\right]\right), \tag{28}$$

and let the λ-optimal discrete-time average value function be given by

$$V_{\lambda,D}^*(x) := \inf_{\pi\in\tilde{\mathscr{P}}} V_{\lambda,D}(x,\pi), \quad x \in S. \tag{29}$$

It follows that $D \mapsto V_{\lambda,D}^*(\cdot)$ is monotone and additively homogeneous, that is,

$$V_{\lambda,D}^*(\cdot) \leq V_{\lambda,D_1}^*(\cdot) \text{ if } D \leq D_1, \quad \text{and} \quad V_{\lambda,c+D}^*(\cdot) = c + V_{\lambda,D}^*(\cdot), \tag{30}$$

where $c \in \mathbb{R}$. Since $D \leq D_1 + \|D - D_1\|$ it follows that $V_{\lambda,D}^*(\cdot) \leq V_{\lambda,D_1}^*(\cdot) + \|D - D_1\|$, and interchanging the role of D and D_1 this yields that

$$\|V_{\lambda,D}^* - V_{\lambda,D_1}^*\| \leq \|D - D_1\|. \tag{31}$$

Using that $V_{\lambda,0}^* = 0$, the monotonicity property in (30) yields that, for $D, D_1 \in \mathbb{B}(S)$,

$$V_{\lambda,D}^* \leq 0 \leq V_{\lambda,D_1}^* \text{ when } D \leq 0 \leq D_1. \tag{32}$$

Theorem 6.1 *Suppose that for each $x, y \in S$, the mapping $a \mapsto p_{x,y}(a)$ is continuous in $a \in A(x)$. Under Assumptions 2.1(i) and 3.1, the following assertions hold:*

(i) For each $D \in \mathbb{B}(S)$ there exist $\mu_D \in \mathbb{R}$ and $h_D : S \to \mathbb{R}$ such that

$$e^{\lambda[\mu_D + h_D(x)]} = \inf_{a\in A(x)} \left[e^{\lambda D(x,a)} \sum_{y\in S} p_{x,y}(a) e^{\lambda h(y)}\right],$$
$$\mu_D = V_{\lambda,D}(x), \quad x \in S. \tag{33}$$

(ii) For each $D, D_1 \in \mathbb{B}(S)$,

$$|\mu_D - \mu_{D_1}| \leq \|D - D_1\|. \tag{34}$$

Part (i) was proved in [10], whereas part (ii) follows from (31) and (33).

Lemma 6.1 *Suppose that Assumption 2.1 is valid, and for each* $g \in \mathbb{R}$ *define the function* $D_g : \mathbb{K} \to \mathbb{R}$ *by*

$$D_g(x,a) = C(x,a) + \frac{1}{\lambda} \log\left(\int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\, dt - gs]}\, dF_{x,a}(s) \right), \quad (x,a) \in \mathbb{K}. \tag{35}$$

With this notation, the following assertions (i)–(v) hold:

(i) $D_g \in \mathbb{B}(\mathbb{K})$ *for each* $g \in \mathbb{R}$.
(ii) $\|D_g - D_{g_1}\| \leq B|g - g_1|, \quad g, g_1 \in \mathbb{R}$.
(iii) There exist $g^- \geq 0$ *such that* $D_{g^-} \leq 0$.
(iv) For each $\beta > 0$, *there exists* $M_\beta > 0$ *such that, when* $\mu \geq M_\beta$, *the inequality* $\int_0^B e^{\mu s}\, dF_{x,a}(s) > \beta$ *holds for every* $(x,a) \in \mathbb{K}$.
(v) $D_{g^+} \geq 0$ *for some* $g^+ \leq 0$.

Proof (i) Let $x \in S$ and $g \in \mathbb{R}$ be arbitrary. As in the proof of Theorem 4.1(ii), the mapping $a \mapsto \int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\, \mathrm{d}t - gs]}\, \mathrm{d}F_{x,a}(s)$ is continuous in its domain $A(x)$, and then the continuity of $C(x,\cdot)$ yields that $D_g(x,\cdot)$ is also continuous. Since the state space is endowed with the discrete topology, it follows that $D_g \in \mathbb{B}(\mathbb{K})$.
(ii) Observing that the inequality $e^{\lambda[\int_0^s \rho_{x,a}(t)\, \mathrm{d}t - gs]} \leq e^{\lambda[\int_0^s \rho_{x,a}(t)\, \mathrm{d}t - g_1 s]} e^{\lambda B|g - g_1|}$ is always valid for every $s \in [0,B]$, the desired conclusion follows via (35).
(iii) Let $\alpha = e^{-\lambda\|C\|}/2$ and select m_α as in Lemma 4.1(ii). Using (7), note that, for every $(x,a) \in \mathbb{K}$, $\int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\, \mathrm{d}t - (B_\rho + \gamma)s]}\, \mathrm{d}F_{x,a}(s) \leq \int_0^B e^{-\lambda\gamma s}\, \mathrm{d}F_{x,a}(s) \leq \alpha$ when $\gamma \geq m_\alpha/\lambda$. Thus, setting $g^- := B_\rho + m_\alpha/\lambda$, via (35) it follows that the inequality $D_{g^-}(x,a) \leq \|C\| + \log(\alpha)/\lambda \leq \|C\| - \|C\| - \log(2)/\lambda < 0$ holds for every $(x,a) \in \mathbb{K}$.
(iv) Let $x \in S$ be arbitrary. By Assumption 2.1, the mapping $r_n(a) \mapsto 1/\int_0^B e^{ns}\, dF_{x,a}(s)$ is continuous in $a \in A(x)$ for every $n \in \mathbb{N}$. Moreover, (1) and the monotone convergence theorem yield that $r_n \searrow 0$, and then, recalling that $A(x)$ is compact, Dini's theorem implies that for each $\beta > 0$ there exists $M_{x,\beta} \in \mathbb{N}$ such that, for every $a \in A(x)$, the inequality $1/\int_0^B e^{ns}\, dF_{x,a}(s) \leq 1/\beta$ holds when $n \geq M_{x,\beta}$, and the conclusion follows setting $M_\beta := \max\{M_{x,\beta} \mid x \in S\}$.
(v) Set $\beta = e^{\lambda\|C\|} > 0$ and select M_β as in part (iv). Using (7), note that, for every $(x,a) \in \mathbb{K}$, $\int_0^B e^{\lambda[\int_0^s \rho_{x,a}(t)\, \mathrm{d}t + (B_\rho + M_\beta/\lambda)s]}\, \mathrm{d}F_{x,a}(s) \geq \int_0^B e^{M_\beta s}\, \mathrm{d}F_{x,a}(s) \geq \beta = e^{\lambda\|C\|}$, a relation that, via (35), yields that $D_{g^+} \geq 0$ for $g^+ := -(B_\rho + M_\beta/\lambda)$. □

Proof of Theorem 3.2 For each $g \in \mathbb{R}$, consider the function $D_g \in \mathbb{B}(\mathbb{K})$ defined in (35). By Theorem 6.1, the discrete-time risk-sensitive average cost $V_{\lambda,D_g}(\cdot)$ is the constant μ_{D_g}, whereas Lemma 6.1(ii) and (34) together yield that $g \mapsto \mu_{D_g}$ is a continuous mapping in $g \in \mathbb{R}$. Now, let g^+ and g^- be as in parts (iii) and (v) of

Lemma 6.1, so that $D_{g^+} \geq 0$ and $D_{g^-} \leq 0$. In this case $\mu_{D_{g^+}} \geq 0$ and $\mu_{D_{g^-}} \leq 0$, by (32) and (33). From this point, the intermediate value property yields the existence of a real number g^* between g^+ and g^- such that $\mu_{D_{g^*}} = 0$, and then Theorem 6.1(i) ensures that for a certain function $h_{D_{g^*}} : S \to \mathbb{R}$ the following equality holds for every $x \in S$:

$$e^{\lambda h_{D_{g^*}}(x)} = \inf_{a \in A(x)} \left[e^{\lambda D_{g^*}(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{D_{g^*}}(y)} \right]$$

$$= \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \int_0^B e^{\lambda \left[\int_0^s \rho_{x,a}(t)\, dt - g^* s \right]} \mathrm{d}F_{x,a}(s) \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{D_{g^*}}(y)} \right],$$

where the second equality is due to (35). Thus, the pair $(g^*, h_{D_{g^*}}(\cdot))$ satisfies the optimality Eq. (16). □

7 Conclusions

In this paper semi-Markov decision chains endowed with the risk-sensitive average criterion were studied. An optimality equation was formulated, and under Assumption 2.1, it was proved that if such an equation has a solution, then a risk-sensitive average optimal stationary policy can be immediately determined, and that the optimal risk-sensitive average cost function is constant. If, additionally, the communication condition in Assumption 3.1 holds, then it was verified that the optimality equation has a solution. As already noted after the statement of Theorem 3.1, in the present risk-sensitive context the right-hand side of the optimality equation involves the *whole* distribution functions of the sojourn times, a fact that establishes an interesting contrast with the risk-neutral case, where the optimality equality engages only the expectations of the sojourn times and the running costs; this difference can be traced back to the nonlinearity of the mapping $Y \mapsto \mathscr{E}_\lambda[Y]$ when $\lambda > 0$. On the other hand, when the communication condition in Assumption 3.1 fails, the optimal risk-sensitive average cost function is not constant in general, and in that case it cannot be characterized in terms of a *single* equation. Thus, it is interesting to look for a characterization of the optimal average cost for semi-Markov decision chains with general transition structure.

Acknowledgments The authors are deeply grateful to the reviewers for helpful suggestions to improve the paper. This work was supported in part by the PSF Organization under Grant No. 015/300/02 and by PRODEP under Grant No. 17332-UAAAN-CA23.

References

1. Stidham Jr., S., Weber, R.R.: A survey of Markov decision models for control of networks of queues. Queueing Syst. **13**, 291–314 (1993)
2. Sennott, L.I.: Stochastic Dynamic Programming and the Control of Queueing Systems. Wiley, New York (1999)
3. Tijms, H.C.: A First Course in Stochastic Models. Wiley, New York (2003)

4. Pinedo, M.: Scheduling: Theory, Algorithms, and Systems. Springer, New York (2008)
5. Hu, Q., Yue, W.: Optimal replacement of a system according to a semi-Markov decision process in a semi-Markov environment. Optim. Method Softw. **18**, 181–196 (2003)
6. Luque-Vásquez, F., Hernández-Lerma, O.: Semi-Markov control models with average costs. Appl. Math. **26**, 315–331 (1999)
7. Baykal-Gürsoy, M.: Semi-Markov Decision Processes. Wiley Encyclopedia of Operations Research and Management Sciences. Wiley, New York (2010)
8. Howard, R.A., Matheson, J.E.: Risk-sensitive Markov decision processes. Manag. Sci. **18**, 356–369 (1972)
9. Cavazos-Cadena, R., Fernández-Gaucherand, E.: Controlled Markov chains with risk-sensitive criteria: average cost, optimality equations, and optimal solutions. Math. Method Oper. Res. **49**, 299–324 (1999)
10. Cavazos-Cadena, R., Fernández-Gaucherand, E.: Risk-sensitive control in communicating average Markov decision chains. In: Dror, M., L'Ecuyer, P., Szidarovsky, F. (eds.) Modelling Uncertainty: An Examination of Stochastic Theory, Methods and Applications, pp. 525–544. Kluwer, Boston (2002)
11. Cavazos-Cadena, R.: Solutions of the average cost optimality equation for finite Markov decision chains: risk-sensitive and risk-neutral criteria. Math. Method Oper. Res. **70**, 541–566 (2009)
12. Hernández-Hernández, D., Marcus, S.I.: Risk sensitive control of Markov processes in countable state space. Syst. Control Lett. **29**, 147–155 (1996)
13. Hernández-Hernández, D., Marcus, S.I.: Existence of risk sensitive optimal stationary polices for controlled Markov processes. Appl. Math. Opt. **40**, 273–285 (1999)
14. Di Masi, G.B., Stettner, L.: Risk-sensitive control of discrete time Markov processes with infinite horizon. SIAM J. Control Optim. **38**, 61–78 (1999)
15. Di Masi, G.B., Stettner, L.: Infinite horizon risk sensitive control of discrete time Markov processes with small risk. SIAM J. Control Optim. **40**, 15–20 (2000)
16. Di Masi, G.B., Stettner, L.: Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. SIAM J. Control Optim. **46**, 231–252 (2007)
17. Jaśkiewicz, A.: Average optimality for risk sensitive control with general state space. Ann. Appl. Probab. **17**, 654–675 (2007)
18. Bäuerle, N., Rieder, U.: Markov Decision Processes with Applications to Finance. Springer, New York (2011)
19. Bäuerle, N., Rieder, U.: More risk-sensitive Markov decision processes. Math. Oper. Res. **39**, 105–120 (2013)
20. Alanís-Durán, A., Cavazos-Cadena, R.: An optimality system for finite average Markov decision chains under risk-aversion. Kybernetika **48**, 83–104 (2012)
21. Cavazos-Cadena, R.: A Poisson equation for the risk-sensitive average cost in semi-Markov chains. Discrete Event Dyn. Syst. (2016). doi:10.1007/s10626-015-0224-z
22. Arapostathis, A., Borkar, V.K., Fernández-Gaucherand, E., Gosh, M.K., Marcus, S.I.: Discrete-time controlled Markov processes with average cost criteria: a survey. SIAM J. Control Optim. **31**, 282–334 (1993)
23. Munkres, J.R.: Topology: A First Course. Prentice-Hall, New York (1974)