Time and Ratio Expected Average Cost Optimality for Semi-Markov Control Processes on Borel Spaces

Fernando Luque-Vásquez and Oscar Vega-Amaya

April 24, 2003

Abstract

We deal with semi-Markov control models with Borel state and control spaces, and unbounded cost functions under the ratio and the time expected average cost criteria. Under suitable growth conditions on the costs and the mean holding times together with stability conditions on the embedded Markov chains, we show the following facts: (i) the ratio and the time average costs coincide in the class of the stationary policies; (ii) there exists an stationary policy which is optimal for both criteria. Moreover, we provide a generalization of the classical Wald's Lemma to semi-Markov processes. These results are obtained combining the existence of solutions of the average cost optimality equation and the Optional Stopping Theorem.

1 Introduction

This paper deals with semi-Markov control models (SMCMs) with Borel state and control spaces, unbounded costs and holding mean times. We consider the two main expected average cost criteria studied in SMCMs, namely, the *ratio expected average cost* (ratio-EAC) *criterion* and the *time expected average cost* (time-EAC) *criterion.* It is well-known that these criteria in general differ even for the case of finite state and control spaces (see [4]). In the present paper under stability conditions and suitable growth assumption on the cost we show that these criteria coincide when the processes are controlled by stationary policies and also that there exists an stationary policy which is optimal for both criteria; thus, the optimality criteria defined by the ratio-EAC criterion and the time-EAC criterion are equivalent. As a by-product we also obtain a generalization of the classical Wald's Lemma in renewal theory to semi-Markov processes which is interesting by itself.

The ratio-EAC criterion has been widely studied in many works (see, e.g. [3], [4], [5], [12], [19], [21], [23], [24],[25] and [29]) but there are only few authors that consider the time-EAC (see, e.g. [4], [15], [17], [23], [24], [25] and [30]) and most of them deal with the countable state space and/or use bounded costs.

The approach we use combine some results on the existence of solutions of the Poisson equation for the ratio-EAC criterion and the Optional Stopping Theorem (see [1] p. 279). A similar approach was previously used in [15] but our analysis is more direct and we do require neither to assume the boundness of the mean holding time function nor to impose a uniform integrability on it as it is done in the mentioned reference.

The remainder of the paper is organized as follows. In Section 2 we introduce the SMCM we will be dealing with and in Section 3 we introduce the performance criteria. The assumptions and main results are stated in Section 4 and the proofs are given in Section 5.

We will use the following notation. A Borel set, say $\mathbb{X}$, of a complete and separable metric space is called a Borel space, and it is endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{X})$. If $\mathbb{X}$ and $\mathbb{Y}$ are Borel spaces, a stochastic kernel on $\mathbb{X}$ given $\mathbb{Y}$ is a function $P(\cdot \mid \cdot)$ such that $P(\cdot \mid y)$ is a probability measure on $\mathbb{X}$ for every $y \in \mathbb{Y}$ and $P(B \mid \cdot)$ is a (Borel-) measurable function on $\mathbb{Y}$ for every $B \in \mathcal{B}(\mathbf{X})$. We denote by $\mathbb{N}$ (respectively $\mathbb{N}_0$) the set of positive (resp. nonnegative) integers; $\mathbb{R}$ (resp., $\mathbb{R}_+$) denotes the set of real (resp., nonnegative) numbers.

2 The model

We consider a *semi-Markov control model* (sMCM) specified by

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, G, C),$$

where:

- The *state space* $\mathbf{X}$ and the *control space* $\mathbf{A}$ are both *Borel spaces*.
- For each $x \in \mathbf{X}$, the subset $A(x)$ of $\mathbf{A}$ is the set of *admissible controls* for the state x. We assume that the *admissible pair* set

  $$\mathbb{K} := \{(x, a) : x \in \mathbf{X}, a \in A(x)\},$$

  is a *Borel subset* of the Cartesian product $\mathbf{X} \times \mathbf{A}$ and also that it contains the graph of a measurable function from $\mathbf{X}$ to $\mathbf{A}$. The latter condition guarantees that the class $\mathbb{F}$ of measurable functions $f : \mathbf{X} \to \mathbf{A}$ satisfying the constraint $f(x) \in A(x)$, for every $x \in \mathbf{X}$, is non-empty.
- The *transition law* $Q(B|x, a)$, with $B \in \mathcal{B}(\mathbf{X})$ and $(x, a) \in \mathbb{K}$, is a stochastic kernel on $\mathbf{X}$ given $\mathbb{K}$.
- The *conditional distribution of holding* (or sojourn) *times* $G(t|x, a, y)$ is a distribution function on $\mathbb{R}_+ := [0, \infty)$ for each fixed $(x, a, y) \in \mathbb{K} \times \mathbf{X}$ and a measurable function on $\mathbb{K} \times \mathbf{X}$ for each fixed time $t \in \mathbb{R}_+$. Let

  $$F(t|x, a) := \int_{\mathbf{X}} G(t|x, a, y) Q(dy|x, a) \quad \forall (x, a) \in \mathbb{K}, t \in \mathbb{R}_+ \tag{1}$$

  the (unconditional) *distribution of holding* (or sojourn) *times*.

- Finally, the measurable function C defined on $\mathbb{K} \times \mathbb{R}_+$ is the *cost function.*

A semi-Markov control model can be thought of as a model of a stochastic system evolving as follows: the system is observed at time $s = 0$ in some state $x_0 = x \in \mathbf{X}$ and it is chosen a control $a_0 = a \in A(x)$; then, the system remains in state $x_0 = x$ for a nonnegative random time δ_1 with distribution function given by $F(\cdot|x, a)$ The cost for operating the system up to any time t prior to δ_1 is given as $C(x, a, t)$. At time δ_1, the system moves to a new state $x_1 = y \in \mathbf{X}$ according to the probability measure $Q(\cdot|x, a)$ and, immediately after the transition occurs, a new control $a_1 = a' \in A(y)$ is chosen, and so forth. Thus, let x_n, a_n and δ_{n+1} be the state of the system immediately after the *nth* transition, the control chosen at that epoch and the holding or sojourn time, respectively. Then, the epoch for the *nth* transition is given by

$$T_n := T_{n-1} + \delta_n \quad n \in \mathbb{N}, \quad \text{and} \quad T_0 := 0, \tag{2}$$

and the number of transitions up to time t is given as

$$N(t) := \sup\{n \geq 0 : T_n \leq t\}, \;\; t \geq 0. \tag{3}$$

Now, for each $n \in \mathbb{N}_0$, define the set of *admissible histories* until the *nth* transition by

$$\mathbf{H}_0 := \mathbf{X}, \quad \mathbf{H}_n := (\mathbb{K} \times \mathbb{R}_+)^n \times \mathbf{X} \quad \text{for } n \in \mathbb{N}.$$

Thus, a *control policy* $\pi = \{\pi_n\}$ is a sequence of stochastic kernels on $\mathbf{A}$ given $\mathbf{H}_n$ satisfying the constraint

$$\pi_n(A(x_n)|h_n) = 1 \quad \forall h_n = (x_0, a_0, \delta_1, ..., x_{n-1}, a_{n-1}, \delta_n, x_n) \in \mathbf{H}_n.$$

A policy $\pi = \{\pi_n\}$ is said to be a (deterministic) *stationary policy* if there exists $f \in \mathbb{F}$ such that $\pi_n(\cdot|h_n)$ is concentrated at $f(x_n)$ for each integer number $n \geq 0$. We denote by Π the class of all policies and, following a usual convention, identify the subclass of stationary policies with $\mathbb{F}$.

As it is well-known, for each policy $\pi \in \Pi$ and initial state $x_0 = x \in \mathbf{X}$, there exists a probability measure P_x^π on the measurable space $(\Omega, \mathcal{F})$ which governs the evolution of the process $\{(x_n, a_n, \delta_{n+1})\}$, where $\Omega := (\mathbf{X} \times \mathbf{A} \times \mathbb{R}_+)^\infty$ and $\mathcal{F}$ is the corresponding product σ-algebra. We denote by E_x^π the expectation operator with respect to the probability measure P_x^π.

Throughout the paper we shall use the following notation. For a measurable function v on $\mathbb{K}$ and $f \in \mathbb{F}$, let

$$v_f(x) := v(x, f(x)) \quad x \in \mathbf{X}. \tag{4}$$

In particular, for the transition law we write

$$Q_f(\cdot|x) := Q(\cdot|x, f(x)) \quad x \in \mathbf{X}. \tag{5}$$

Note that, for an arbitrary policy $\pi \in \Pi$, the distribution of the variable state x_n may depend on the evolution of the process during the first $n-1$ transitions. However, under a stationary policy $f \in \mathbb{F}$, the process $\{(x_n, T_n)\}$ is a Markov renewal process with semi-Markov kernel $P(x, B, t) = \int_B G(t \mid x, a, y) Q_f(dy \mid x)$ and the state process $\{x_n\}$ is a (homogeneous) Markov chain with transition probability $Q_f(\cdot|\cdot)$. For the latter case, we denote by $Q_f^n(\cdot|\cdot)$ the n-step transition probability.

The reader can found an nice introduction to (noncontrolled) Markov renewal processes and semi-Markov processes in [18].

3 Expected average cost criteria

Now we introduce the criteria which we are interested in, namely, the so-called *ratio expected average cost* and the *time expected average cost.* Thus, for each policy $\pi \in \Pi$ and initial state $x_0 = x \in \mathbf{X}$, let

$$J_t(\pi, x) := E_x^\pi \sum_{k=0}^{N(t)} C(x_k, a_k, \delta_{k+1}), \tag{6}$$

be the *expected cost up to time* $t > 0$. Then, the *time expected average cost* (time EAC) is defined by

$$J(\pi, x) := \limsup_{t \to \infty} \frac{1}{t} J_t(\pi, x). \tag{7}$$

Now, for each $n \in \mathbb{N}$, let

$$\Phi_n(\pi, x) := E_x^\pi \sum_{k=0}^{n-1} C(x_k, a_k, \delta_{k+1}) \tag{8}$$

be the *expected cost up to the nth transition*; thus, the *ratio expected average cost* (ratio EAC) is given as

$$\Phi(\pi, x) := \limsup_{n \to \infty} \frac{1}{E_x^\pi T_n} \Phi_n(\pi, x). \tag{9}$$

A policy $\pi^* \in \Pi$ is said to be *time expected average cost* (time EAC-) *optimal* if

$$J(\pi, x) \geq J(\pi^*, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi.$$

Similarly, a policy π_* is said to be *ratio expected average cost* (ratio EAC-) *optimal* if

$$\Phi(\pi, x) \geq \Phi(\pi_*, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi.$$

The finite horizon expected costs (6) and (8), as well as the infinite expected average costs (7) and (9), can be expressed in a slightly different but useful way

by considering the *mean holding times*

$$\tau(x,a) := \int_0^\infty sF(ds|x,a) \quad (x,a) \in \mathbb{K}, \tag{10}$$

and the *mean costs*

$$\widehat{C}(x,a) := \int_0^\infty C(x,a,s)F(ds|x,a) \quad (x,a) \in \mathbb{K}. \tag{11}$$

Thus we have

$$J(\pi,x) = \limsup_{t\to\infty} \frac{1}{t} E_x^\pi \sum_{k=0}^{N(t)} \widehat{C}(x_k,a_k), \tag{12}$$

$$\Phi(\pi,x) = \limsup_{n\to\infty} \frac{E_x^\pi \sum_{k=0}^{n-1} \widehat{C}(x_k,a_k)}{E_x^\pi \sum_{k=0}^{n-1} \tau(x_k,a_k)}, \tag{13}$$

for all policy $\pi \in \Pi$ and all initial state $x \in \mathbf{X}$. In fact, one can verify using conditional expectation properties that

$$\Phi_n(\pi,x) := E_x^\pi \sum_{k=0}^{n-1} \widehat{C}(x_k,a_k) \quad \text{and} \quad E_x^\pi T_n = E_x^\pi \sum_{k=0}^{n-1} \tau(x_k,a_k)$$

which yields (13) provided that all the expectations involved are well defined. Now, to obtain (12), observe that

$$\begin{aligned} J_t(\pi,x) &= \sum_{k=0}^\infty E_x^\pi C(x_k,a_k,\delta_{k+1}) \mathbf{I}_{[T_k\le t]} \quad \forall x \in \mathbf{X}, \pi \in \Pi, t \in \mathbb{R}_+ \\ &= \sum_{k=0}^\infty E_x^\pi \left\{ \mathbf{I}_{[T_k\le t]} E_x^\pi \left[C(x_k,a_k,\delta_{k+1}) \,|h_k,a_k\,] \right\} \\ &= \sum_{k=0}^\infty E_x^\pi \widehat{C}(x_k,a_k) \mathbf{I}_{[T_k\le t]}, \end{aligned}$$

where $h_k = (x_0,a_0,\delta_1,...,x_{k-1},a_{k-1},\delta_k,x_k)$, with $x_0 = x$ and $k \in \mathbb{N}_0$. Thus, we obtain

$$J_t(\pi,x) = E_x^\pi \sum_{k=0}^{N(t)} \widehat{C}(x_k,a_k) \quad \forall x \in \mathbf{X}, \pi \in \Pi, t \in \mathbb{R}_+, \tag{14}$$

from which (12) follows.

Finally, it is worth mentioning that in general the time and the ratio EAC criteria differ (see [4], Example 3.1). For the denumerable state space case,

Ross [23] shows that these criteria coincide for stationary policies under a recurrence/ergodic condition; Jaskiewicz in [15] obtains the same result for unbounded costs and Borel spaces under a stability hypothesis similar to the used in the present paper, but she additionally suppose that the mean holding time function is bounded and impose on it a uniform integrability condition.

4 Assumptions and Main Results

We begin imposing a growth condition on both the mean cost and the mean holding time functions introduced in (10) and (11), respectively.

Assumption 4.1. There exist a measurable function $W(\cdot) \geq 1$ on $\mathbf{X}$ and a constant $M > 0$ such that

$$\max\{|\widehat{C}(x,a)|, \tau(x,a)\} \leq MW(x) \quad \forall (x,a) \in \mathbb{K}.$$

Throughout the paper we shall use the following notation: for a measurable function $u(\cdot)$ on $\mathbf{X}$ define the *W-weighted norm*, W-norm for short, as

$$||u||_W := \sup_{x \in \mathbf{X}} \frac{|u(x)|}{W(x)}$$

and denote by $B_W(\mathbf{X})$ the Banach space of measurable functions with finite W-norm. Moreover, for a function $u(\cdot)$ and a measure $\mu(\cdot)$ on $\mathbf{X}$, let

$$\mu(u) := \int_{\mathbf{X}} u(y)\mu(dy)$$

whenever the integral is well-defined.

Now we introduce a second set of conditions which ensure that the "embedded" Markov chains induced by the stationary policies are well behaved in the long-term.

Assumption 4.2. There exist a non-trivial measure $\nu(\cdot)$ on $\mathbf{X}$, a nonnegative measurable function $\phi(\cdot,\cdot)$ on $\mathbb{K}$ and a positive constant $\beta < 1$ such that:

(a) $\nu(W) < \infty$;

(b) $Q(B|x,a) \geq \nu(B)\phi(x,a) \quad \forall B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbb{K}$;

(c) $\int_{\mathbf{X}} W(y)Q(dy|x,a) \leq \beta W(x) + \phi(x,a)\nu(W) \quad \forall (x,a) \in \mathbb{K}$;

(d) For each $f \in \mathbb{F}$ there exists a state $x_f \in \mathbf{X}$ such that $\phi_f(x_f) \neq 0$.

Several variants of Assumption 4.2 have been already considered in a number of papers to study controlled Markov and semi-Makov processes ([7], [10], [11], [12], [21], [29]) as well as Markov and semi-Markov games ([9], [13], [14], [16], [20]). In fact, the stability conditions in Assumption 4.2 are the same that the used in [6], [26], and [27], except for the fact that the latter references suppose, instead of Assumption 4.2(d), that $\nu(\phi_f) > 0$ for all $f \in \mathbb{F}$, which

is obviously a stronger condition. The next remark collects some of the most relevant consequences of Assumption 4.2 with regards of the present work, which can be proved as in [26].

Remark 4.3. Suppose that Assumptions 4.1 and 4.2 hold and let f be a fixed but arbitrary policy in $\mathbb{F}$. Then:

(a) The transition law $Q_f(\cdot|\cdot)$ is *positive Harris recurrent*; hence, in particular, it admits a *unique invariant probability* measure $\mu_f(\cdot)$, that is,

$$\mu_f(B) = \int_{\mathbf{X}} Q_f(B|x)\mu_f(dx) \quad \forall B \in \mathcal{B}(\mathbf{X}). \tag{15}$$

Moreover, $\mu_f(W) < \infty$; thus, $\mu_f(|u|) < \infty$ for all function $u(\cdot)$ in $B_W(\mathbf{X})$.

(b) For every function $u(\cdot)$ in $B_W(\mathbf{X})$, we have

$$\lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} u(x_k) = \mu_f(u) \quad \forall x \in \mathbf{X}; \tag{16}$$

thus, in particular,

$$\lim_{n\to\infty} \frac{1}{n} E_x^f u(x_n) = 0 \quad \forall x \in \mathbf{X}. \tag{17}$$

(c) For each function $u(\cdot)$ in $B_W(\mathbf{X})$, there exists a function $h_u(\cdot)$ in $B_W(\mathbf{X})$ that solves the Poisson equation

$$h_u(x) = u(x) - \mu_f(u) + \int_{\mathbf{X}} h_u(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}. \tag{18}$$

It is important to mention, on one hand, that Assumption 4.2 as well as the variants used in the previously cited papers yields the property known as *W-geometric ergodicity* (see, for instance, [8], [9] and [22]): for each $f \in \mathbb{F}$, there exist positive constants $\lambda_f < 1$ and $M_f < \infty$ such that

$$||Q_f^n u - \mu_f(u)||_W \leq M_f \lambda_f^n \quad \forall u \in B_W(\mathbf{X}), n \in \mathbb{N}.$$

Using this property it is easy to check that the function

$$h(x) := \sum_{k=0}^{\infty} [Q_f^n u(x) - \mu_f(u)] \quad \forall x \in \mathbf{X},$$

is in $B_W(\mathbf{X})$ and that it satisfies the Poisson equation (18). On the other hand, it turns out that Assumption 4.2 is a contraction property, which is exploited in [26] to give pretty direct proofs of all results in Remark 4.3 using solely fixed-point arguments.

As it is well-known the relevance of the Poisson equation comes from the fact that it allows the analysis of the accumulated cost up to the time of the *nth* transition, specially when n is large enough or tends to infinity. To illustrate this fact suppose that

$$\tau(x,a) > 0 \quad \forall (x,a) \in \mathbb{K}, \tag{19}$$

and define

$$\rho_f := \frac{\mu_f(\widehat{C}_f)}{\mu_f(\tau_f)} \quad \forall f \in \mathbb{F}. \tag{20}$$

Note that under Assumptions 4.1 and 4.2, these constants are finite because of Remark 4.3(a). Thus, for a stationary policy $f \in \mathbb{F}$, letting $u(\cdot) = \widehat{C}_f(\cdot) - \rho_f \tau_f(\cdot)$, from Remark 4.3(c) we have a function $h_f(\cdot)$ in $B_W(\mathbf{X})$ satisfying the Poisson equation

$$h_f(x) = \widehat{C}_f(x) - \rho_f \tau_f(x) + \int_{\mathbf{X}} h_f(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}. \tag{21}$$

Then, iterations yield

$$\Phi_n(f,x) = \rho_f E_x^f \sum_{k=0}^{n-1} \tau_f(x_k) + h_f(x) - E_x^f h_f(x_n) \quad \forall x \in \mathbf{X}, n \in \mathbb{N}. \tag{22}$$

Hence, as a direct consequence of Remark 4.3(b) and (19), we have

$$\Phi(f,x) = \rho_f \quad \forall x \in \mathbf{X}. \tag{23}$$

Proceeding as above, for each $f \in \mathbb{F}$, we also have a function $\widehat{h}_f \in B_W(\mathbf{X})$, such that

$$\widehat{h}_f(x) = \widehat{C}_f(x) - \mu_f(\widehat{C}_f) + \int_{\mathbf{X}} \widehat{h}_f(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}, \tag{24}$$

which implies

$$E_x^f \sum_{k=0}^{n-1} \widehat{C}_f(x_k) = n\mu_f(\widehat{C}_f) + \widehat{h}_f(x) - E_x^f \widehat{h}_f(x_n) \quad \forall x \in \mathbf{X}, n \in \mathbb{N}. \tag{25}$$

Now, to get a continuous-time analogous of (22) and (25) we need to reinforce the condition (19) as follows:

Assumption 4.4. There exist positive constants σ and ϵ such that

$$F(\sigma|x,a) \leq 1 - \epsilon \quad \forall (x,a) \in \mathbb{K}.$$

The condition in Assumption 4.4 is mainly used in the related literature to guarantee that the regularity property holds, that is, to preclude the semi-Markov processes experience infinitely many transitions on any bounded interval of time. However, it is worth mentioning that under Assumption 4.2 and condition (19) we have of course the regularity property for the semi-Markov processes

induced by stationary policies. In fact, the paper [28] shows that the regularity property holds under the weaker condition of Harris recurrence. Thus, in the present paper, the role of Assumptions 4.4 is to assure the stronger properties stated in the first part of the next lemma.

Lemma 4.5. If Assumption 4.4 holds, then:

(a) $E_x^\pi N(t) \leq \dfrac{t+\sigma}{\varepsilon\sigma} \quad \forall x \in \mathbf{X}, \pi \in \Pi, t > 0;$

(b) $\tau(x,a) \geq \varepsilon\sigma \quad \forall (x,a) \in \mathbb{K}.$

The first part of Lemma 4.5 comes from [30], whereas the second one can be easily obtained using the integration by parts rule.

Theorem 4.6. Suppose that Assumptions 4.1, 4.2 and 4.4 hold and let $h_f(\cdot)$ and $\widehat{h}_f(\cdot)$ solutions of the Poisson equations (21) and (24), respectively. Then, for each $f \in \mathbb{F}$:

$$J_t(f,x) = \rho_f E_x^f \sum_{k=0}^{N(t)} \tau_f(x_k) + h_f(x) - E_x^f h_f(x_{N(t)+1}), \tag{26}$$

and

$$J_t(f,x) = \mu_f(\widehat{C}_f) E_x^f[N(t)+1] + \widehat{h}_f(x) - E_x^f \widehat{h}_f(x_{N(t)+1}) \tag{27}$$

for all $x \in \mathbf{X}$ and $t > 0$.

Notice that both expressions (26) and (27) give ways for analyzing the long-term behavior of the costs in continuous time, but the latter one has the appealing that it can be thought of as a generalization of the Wald's Lemma in the classical renewal theory. In words, it states that the expected cost up to time $t > 0$ equals the steady-state cost times the expected number of transitions up to time t plus some terms correcting the deviation from the steady-state regime.

Theorem 4.7. Suppose that Assumptions 4.1, 4.2 and 4.4 hold. Then

$$J(f,x) = \Phi(f,x) = \rho_f \quad \forall x \in \mathbf{X}, f \in \mathbb{F}.$$

Our next main result, Theorem 4.9, states two facts: the equality of the optimal values of the ratio and the time expected average costs and the existence of an stationary optimal policy for both criteria. Now, the standard way to prove the existence of ratio EAC-optimal stationary policies is provided by the so-called *Average Cost Optimality Equation* (see Proposition 5.4, Section 5), which can be thought as a "generalization" of the Poisson equation (21). However, to get this extension is needed to impose some continuity/compactness conditions on the model.

Assumption 4.8. For each state $x \in \mathbf{X}$:

(a) $A(x)$ is a compact subset of $\mathbf{A}$;

(b) $\widehat{C}(x,\cdot)$ is lower semicontinuous on $A(x)$;

(c) $\tau(x,\cdot)$ and $\phi(x,\cdot)$ are continuous functions on $A(x)$;

(d) the transition law $Q(\cdot|x,\cdot)$ is strongly continuous on $A(x)$, that is, the mapping

$$a \longmapsto \int_{\mathbf{X}} u(y)Q(dy|x,a)$$

is continuous for each bounded measurable function u on $\mathbf{X}$. Additionally, the above condition holds with $u = W$.

Theorem 4.9. Suppose that Assumptions 4.1, 4.2, 4.4 and 4.8 hold. Then there exists a triplet (ρ^*, f^*, h^*), where ρ^* is a constant, $f^* \in \mathbb{F}$, and $h^* \in B_W(\mathbf{X})$ such that

$$h^*(x) = \inf_{\pi\in\Pi}\left[J_t(\pi,x) - \rho^* E_x^\pi \sum_{k=0}^{N(t)} \tau(x_k,a_k) + E_x^\pi h^*(x_{N(t)+1})\right]$$

$$= J_t(f^*,x) - \rho^* E_x^{f^*} \sum_{k=0}^{N(t)} \tau_{f^*}(x_k) + E_x^{f^*} h^*(x_{N(t)+1}).$$

for all $x \in \mathbf{X}$ and $t > 0$. Hence, ρ^* is the *time EAC-optimal value* and f^* is *time EAC-optimal*. Thus, from Proposition 4.9,

$$J(f^*,x) = \Phi(f^*,x) = \rho^* = \inf_{\pi\in\Pi} J(\pi,x) = \inf_{\pi\in\Pi}\Phi(\pi,x) \quad \forall x \in \mathbf{X}.$$

5 Proofs

Lemma 5.1. Suppose that Assumptions 4.1, 4.2 and 4.4 hold. Then, there is a constant L such that

$$E_x^\pi \sum_{k=0}^{N(t)} \left|\widehat{C}(x_k,a_k)\right| \le LW(x)\frac{t+\sigma}{\varepsilon\sigma}, \tag{28}$$

and

$$E_x^\pi \sum_{k=0}^{N(t)} \tau(x_k,a_k) \le LW(x)\frac{t+\sigma}{\varepsilon\sigma}, \tag{29}$$

hold for all policy $\pi \in \Pi$, initial state $x \in \mathbf{X}$ and time $t > 0$.

Proof of Lemma 5.1. Observe, by Assumption 4.2(c), that

$$E_x^\pi[W(x_n) \mid x_0,\delta_1,...,\delta_k] \le [1 + \frac{\nu(W)}{1-\beta}]W(x) \quad \forall n,k \in \mathbb{N}, x \in \mathbf{X}. \tag{30}$$

Then, letting $L' := [1 + \nu(W)/(1-\beta)]$, we have

$$\begin{aligned} E_x^\pi \sum_{k=0}^{N(t)} \left|\widehat{C}(x_k, a_k)\right| &= \sum_{k=0}^{\infty} E_x^\pi \left|\widehat{C}(x_k, a_k)\right| \mathbf{I}_{[T_k \le t]} \le M \sum_{k=0}^{\infty} E_x^\pi [W(x_k) \mathbf{I}_{[T_k \le t]}] \\ &= M \sum_{k=0}^{\infty} E_x^\pi [\mathbf{I}_{[T_k \le t]} E_x^\pi [W(x_k) \mid x_0, \delta_1, ..., \delta_k]] \\ &\le ML'W(x) E_x^\pi \sum_{k=0}^{\infty} \mathbf{I}_{[T_k \le t]} \\ &= LW(x) E_x^\pi N(t) \le LW(x) \frac{t+\sigma}{\varepsilon\sigma}, \end{aligned}$$

with $L = ML'$. In a similar way we can prove (29).■

Lemma 5.2. Suppose that Assumptions 4.1, 4.2 and 4.4 hold. Then, for each $x \in \mathbf{X}$ and $\pi \in \Pi$,

$$\lim_{t\to\infty} \frac{1}{t} E_x^\pi W(x_{N(t)+1}) = 0. \tag{31}$$

Proof of Lemma 5.2. For each $t > 0$,

$$\begin{aligned} E_x^\pi W(x_{N(t)+1}) &= E_x^\pi \sum_{k=1}^{\infty} W(x_k) \mathbf{I}_{[T_{k-1} \le t < T_k]} \\ &= \sum_{k=1}^{\infty} E_x^\pi [\mathbf{I}_{[T_{k-1} \le t < T_k]} E_x^\pi [W(x_k) \mid x_0, \delta_1, ..., \delta_k]] \\ &\le L'W(x) \sum_{k=1}^{\infty} P_x^\pi [N(t) = k-1] = L'W(x). \end{aligned}$$

from which (31) follows.■

An immediate consequence of Lemma 5.2 is that for each $u \in B_W(\mathbf{X})$,

$$\lim_{t\to\infty} \frac{1}{t} E_x^\pi u(x_{N(t)+1}) = 0 \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{32}$$

Analogously we can show that

$$\lim_{t\to\infty} \frac{1}{t} E_x^\pi u(x_{N(t)}) = 0. \tag{33}$$

Proof of Theorem 4.6. Let f be a fixed but arbitrary stationary policy and let $h_f(\cdot)$ be a solution of the Poisson equation

$$h_f(x) = \widehat{C}_f(x) - \rho_f \tau_f(x) + \int_{\mathbf{X}} h_f(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}.$$

Next, introduce the process

$$M_n := \sum_{k=0}^{n-1} \widehat{C}_f(x_k) - \rho_f \sum_{k=0}^{n-1} \tau_f(x_k) + h_f(x_n) - h_f(x_0) \quad \text{for } n \geq 1, \text{ and } \quad M_0 := 0,$$

and observe that $\{M_n\}$ is a martingale with respect to the filtration

$$\mathcal{F}_n := \sigma(x_0, a_0, \delta_1, ..., x_{n-1}, a_{n-1}, \delta_n, x_n) \quad \forall n \geq 0.$$

Thus, noting that for $t > 0$ the random variable $N(t) + 1$ is a stopping time with respect to $\{\mathcal{F}_n\}$, and using (28) and (29) we have that

$$E_x^f \left| M_{N(t)+1} \right| < \infty \quad \forall x \in \mathbf{X}, t > 0,$$

and also that

$$\lim_{n \to 0} E_x^f \left| M_n \right| \mathbf{I}_{[N(t) \geq n]} = 0 \quad \forall x \in \mathbf{X}, t > 0.$$

Thus, by the Optional Stopping Theorem, we obtain

$$E_x^f M_{N(t)+1} = E_x^f \sum_{k=0}^{N(t)} [\widehat{C}_f(x_k) - \rho_f \tau_f(x_k) + h_f(x_{N(t)+1}) - h_f(x_0)]$$

$$= E_x^f M_1 = 0.$$

Hence,

$$J_t(f, x) = \rho_f E_x^f \sum_{k=0}^{N(t)} \tau_f(x_k) + h_f(x) - E_x^f h_f(x_{N(t)+1}) \quad \forall x \in \mathbf{X}, t > 0,$$

which proves the first statement of the theorem.

The proof of the second part follows the same arguments but now considering a function $\widehat{h}_f(\cdot)$ in $B_W(\mathbf{X})$ that solves the Poisson equation

$$\widehat{h}_f(x) = \widehat{C}_f(x) - \mu_f(\widehat{C}_f) + \int_{\mathbf{X}} \widehat{h}_f(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}. \blacksquare$$

Lemma 5.3. Suppose that Assumptions 4.1, 4.2 and 4.4 hold. Then,

$$\lim_{t \to \infty} \frac{1}{t} E_x^f [N(t) + 1] = \frac{1}{\mu_f(\tau_f)} \quad \forall x \in \mathbf{X}, f \in \mathbb{F}.$$

Proof of Lemma 5.3. Fix $f \in \mathbb{F}$ and let $w_f \in B_W(\mathbf{X})$ a function satisfying the Poisson equation

$$w_f(x) = \tau_f(x) - \mu_f(\tau_f) + \int w_f(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}.$$

Thus, as in the proof of Theorem 4.6, the Optional Stopping Theorem yields

$$E_x^f \sum_{k=0}^{N(t)} \tau_f(x_k) = \mu_f(\tau_f) E_x^f[N(t)+1] + w(x) - E_x^f w(x_{N(t)+1}), \tag{34}$$

for all $x \in \mathbf{X}$ and $t > 0$. Also, by using properties of the conditional expectation one can show that

$$E_x^f T_{N(t)} = E_x^f \sum_{k=0}^{N(t)-1} \tau_f(x_k) \quad \text{and} \quad E_x^f T_{N(t)+1} = E_x^f \sum_{k=0}^{N(t)} \tau_f(x_k), \tag{35}$$

holds for all $x \in \mathbf{X}, t > 0$. Thus by (34) we have

$$1 \leq \frac{1}{t} E_x^f T_{N(t)+1} = \mu_f(\tau_f) \frac{1}{t} E_x^f[N(t)+1] - \frac{1}{t} E_x^f w(x_{N(t)+1}) + \frac{1}{t} w(x),$$

which by (32) implies

$$\liminf_{t\to\infty} \frac{1}{t} E_x^f[N(t)+1] \geq \frac{1}{\mu_f(\tau_f)}$$

On the other hand, by (35) we see that

$$\frac{1}{t} E_x^f T_{N(t)+1} \leq 1 + \frac{1}{t} E_x^f \tau_f(x_{N(t)}),$$

which combined with (34) yields

$$\mu_f(\tau_f) \frac{1}{t} E_x^f[N(t)+1] - \frac{1}{t} E_x^f w(x_{N(t)+1}) + \frac{1}{t} w(x) \leq 1 + \frac{1}{t} E_x^f \tau(x_{N(t)}).$$

Hence,

$$\limsup_{t\to\infty} \frac{1}{t} E_x^f[N(t)+1] \leq \frac{1}{\mu_f(\tau_f)} \quad \forall x \in \mathbf{X}, t > 0,$$

which proves the desired result.■

Proof of Theorem 4.7. This is a direct consequence of (27), (32) and Lemma 5.3.■

The results in the next proposition can be proved as in [27]; alternative approaches can be founded in several other papers which consider similar but different conditions (see, for instance, [12], [21], [29]).

Proposition 5.4. Suppose Assumptions 4.1, 4.2 and 4.8 hold. If, additionally, $\tau(x,a) > 0$ for all (x,a) in $\mathbb{K}$, then there exists a triplet formed by a constant

ρ^*, an stationary policy $f^* \in \mathbb{F}$, and a function $h^* \in B_W(\mathbf{X})$ that solves the ACOE, that is,

$$\begin{aligned} h^*(x) &= \min_{a \in A(x)} \left[\widehat{C}(x,a) - \rho^* \tau(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \right] \\ &= \widehat{C}_{f^*}(x) - \rho^* \tau_{f^*}(x) + \int_{\mathbf{X}} h^*(y) Q_{f^*}(dy|x), \end{aligned}$$

for all $x \in \mathbf{X}$. Hence, f^* is *ratio EAC-optimal* and ρ^* is *the ratio EAC optimal value*, that is,

$$\Phi(f^*, x) = \inf_{\pi \in \Pi} \Phi(\pi, x) = \rho^* \quad \forall x \in \mathbf{X}.$$

Now we finally proceed to prove Theorem 4.9.

Proof of Theorem 4.9. Let the triplet (f^*, h^*, ρ^*) be as in Proposition 5.4. Fix an arbitrary policy $\pi \in \Pi$ and an arbitrary state $x \in \mathbf{X}$. Next define $U_0 := 0$ and

$$U_n := \sum_{k=0}^{n-1} [\widehat{C}(x_k, a_k) - \rho^* \tau(x_k, a_k)] + h^*(x_n) - h^*(x_0), \quad n = 1, 2, ...$$

It is easy to show that $\{U_n\}$ is a submartingale with respect to the filtration $\{\mathcal{F}_n\}$; moreover, by (28), (29) and (30), it follows that

$$E_x^\pi \left| U_{N(t)+1} \right| < \infty,$$

and

$$\lim_{n \to \infty} E_x^\pi \left| U_n \right| \mathbf{I}_{[N(t) \geq n]} = 0.$$

Then, using again the Optional Stopping Theorem we have

$$E_x^\pi \left[\sum_{k=0}^{N(t)} [\widehat{C}(x_k, a_k) - \rho^* \tau(x_k, a_k)] + h^*(x_{N(t)+1}) - h^*(x_0) \right] \geq E_x^\pi U_1 \geq 0 \quad (36)$$

which combined with (35) yields

$$\frac{1}{t} J_t(\pi, x) + \frac{1}{t} E_x^\pi h^*(x_{N(t)+1}) - \frac{1}{t} h^*(x) \geq \rho^* \frac{1}{t} E_x^\pi T_{N(t)+1} \geq \rho^*.$$

Thus by (32) we have

$$J(\pi, x) \geq \rho^*. \quad (37)$$

Now observe that if we take $\pi = f^*$ the process $\{U_n\}$ is martingale. So all inequalities in (36) and (37) become equalities, proving thus the desire result.■

Acknowledgment. The authors thank to A. Jáskiewicz who kindly showed them her unpublished paper.

References

[1] R.B. Ash and C.A. Doléans-Dade (2000), *Probability and Measure Theory*, Academic Press.

[2] S. Bhatnagar and V.S. Borkar (1995), *A convex analytic framework for ergodic control of semi-Markov processes,* Math. Oper. Res. 20, 923-936.

[3] R.N. Bhattacharya and M. Majumdar (1989), *Controlled semi-Markov models under long-run average rewards*, J. Statist. Plann. Inference 22, 223-242.

[4] E.A. Feinberg (1994), *Constrained semi-Markov decision processes with average rewards,* ZOR-Math. Methods. Oper. Res. 39, pp. 257-267.

[5] A. Federgruen, P.J. Schweitzer and H.C. Tijms (1983), *Denumerable undiscounted semi-Markov decision processes with unbounded rewards,* Math. Oper. Res. 8, 298-213.

[6] J.I. González-Trejo, O. Hernández-Lerma, LF Hoyos-Reyes (2003), *Minimax control of discrete-time stochastic systems*, SIAM J. Control Optim. 41, 1626-1659.

[7] E. Gordienko and O. Hernández-Lerma (1995), *Average cost Markov control processes with weighted norms: existence of canonical policies*, Appl. Math. (Warsaw) 23, 199-218.

[8] O. Hernández-Lerma and J.B. Lasserre (1999), *Further Topics on Discrete-Time Markov Control Processes*, Springer-Verlag, NY.

[9] O. Hernández-Lerma and J.B. Lasserre (2001), *Zero-sum stochastic games in Borel spaces: average payoff*, SIAM J. Control Optim. 39, 1520-1539.

[10] O. Hernández-Lerma and O. Vega-Amaya (1998), *Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality*, Appl. Math. (Warsaw) 25, 153-178.

[11] O. Hernández-Lerma, O. Vega-Amaya and G. Carrasco (1998), *Sample-path optimality and variance-minimization of average cost Markov control processes*, SIAM J. Control and Optim.38, 79-93.

[12] A. Jaśkiewicz (2001), *An approximation approach to ergodic semi-Markov control processes*, Math. Methods Opr. Res. 54, 1-19.

[13] A. Jaśkiewicz (2002), *Zero-sum semi-Markov games*, SIAM J. Control Optim. 41, 723-739.

[14] A. Jaśkiewicz and A.S. Nowak (2001), *On the optimality equation for zero-sum ergodic stochastic games*, Math. Methods Opr. Res. 54, 291-301.

[15] A. Jaśkiewicz (), *On the equivalence of two expected average cost criteria for semi-Markov control processes*

[16] H.-U. Küenle, R. Schurath (2003), *The optimality equation and ε–optimal strategies in Markov games with average reward criterion*, Math. Methods Oper. Res. 56, 439-449.

[17] M. Kurano (1985), *Semi-Markov decision processes and their applications in replacement models*, J. Oper. Res. Soc. Japan 28, 18-29.

[18] N. Limnios and G. Oprişan (2001), *Semi-Markov Processes and Reliability*, Birkhäuser, Boston.

[19] S.A. Lippman (1975), *On dynamic programming with unbounded rewards,* Manage. Sci. 21, 1225-1233.

[20] F. Luque-Vásquez (1992), *Zero-sum semi-Markov games in Borel spaces: discounted and average payoff,* Bol. Soc. Mat. Mexicana 8, 227-241.

[21] F. Luque-Vásquez and O. Hernández-Lerma (1999), *Semi-Markov control models with average costs,* Appl. Math. (Warsaw) 26, 315-331.

[22] S.P. Meyn and R.L. Tweedie (1993), *Markov Chain and Stochastic Stability*, Springer-Verlag, London.

[23] S.M. Ross (1970), *Applied Probability Models with Optimization Applications*, Holden Day, San Francisco.

[24] M. Schäl (1992), *On the second optimality equation for semi-Markov decision models,* Math. Oper. Res. 17, 470-486.

[25] L.I. Sennott (1989), *Average cost semi-Markov decision processes and the control of queueing systems*, Prob. Eng. Inform. Sci. 3, 247-272.

[26] O. Vega-Amaya (2001), *The Average cost optimality equation: a fixed point approach*, To appear in Bol. Soc. Mat. Mexicana. Available in : http://fractus.mat.uson.mx/˜tedi/reportes.

[27] O. Vega-Amaya (2002), *Zero-sum semi-Markov games: Fixed point solutions of the Shapley equation*, To appear in SIAM J. Control Optim. Available in : http://fractus.mat.uson.mx/˜tedi/reportes.

[28] O. Vega-Amaya (2002), *A note on the regularity property of semi-Markov processes with Borel state space*, To appear in Stat. Prob. Letters. Available in : http://fractus.mat.uson.mx/˜tedi/reportes.

[29] O. Vega-Amaya and F. Luque-Vásquez (2000), *Sample-path average cost optimality for semi-Markov control processes on Borel spaces: Unbounded costs and mean holding times*, Appl. Math 27, 343-367.

[30] A.A. Yushkevich (1981), *On semi-markov controlled models with an average reward criterion*, Theory of Probability and its Applications, XXVI, 796-803.