CONTROLLED MARKOV CHAINS WITH RISK-SENSITIVE CRITERIA: SOME (COUNTER) EXAMPLES

Agustin Brau-Rojas
Department of Mathematics, The University of Arizona,
Rolando Cavazos-Cadena
Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, México,
Emmanuel Fernández-Gaucherand
Systems and Industrial Engineering Department, The University of Arizona

Abstract

This paper is concerned with risk-sensitive versions of the standard average cost and discounted cost criteria for controlled Markov chains. Risk-sensitivity is modelled by means of the family of exponential disutility functions. First, we study the risk-sensitive average cost corresponding to a fixed stationary deterministic policy, for finite state space models. We examine some (counter) examples which illustrate how the behavior of the risk-sensitive model departs from that of the risk-null one in the non-irreducible case. Finally, we present an example of a controlled Markov chain with infinite state space which, as opposed to Jaquette’s result for finite models, does not have ultimately stationary optimal policies with respect to the risk-sensitive (exponential) discounted cost. The controlled Markov chain in that example satisfies a simultaneous Doeblin condition, an assumption that enables the vanishing discount approach in the risk-null models. Thus, the example provides more evidence to think that obtaining the mentioned approach for the risk-sensitive models might be even harder than what Jaquette’s result had already indicated.

1 Introduction

In this paper we investigate the risk-sensitive versions of the standard average cost and the discounted cost criteria for controlled Markov chains (CMC’s) introduced by Howard and Matheson [14] and Jaquette [15], respectively; see [4, 7, 6, 5, 9, 10, 11, 16] for other recent results on these criteria. Risk-sensitivity is modelled in these criteria by means of an exponential disutility function \( U_x(x) = \text{sgn}(\gamma) e^{\gamma x} \), where \( \gamma \neq 0 \) is the constant risk-sensitivity coefficient, so that we will refer to them as the exponential average cost (EAC) and the exponential discounted cost (EDC), respectively; see [16, 18, 20]. The paper is organized as follows. In Section 2, the model and some general notation and terminology are introduced. In Section 3, we discuss the EAC corresponding to a fixed (but otherwise arbitrary) stationary deterministic policy \( f^{\infty} \). First, we state a result, Theorem 1, which characterizes the EAC in terms of the spectral radii of certain matrices associated to the irreducible communicating classes of both recurrent and transient states. Theorem 1 generalizes Howard-Matheson’s characterization, which treats only the particular case in which the probability transition matrix \( P \) induced by \( f \) is primitive. Then, we employ Theorem 1 to illustrate, by means of two examples, two peculiar features of the EAC when the initial state is transient: 1) The EAC is not affected by the probabilities of entering the irreducible closed classes the initial state leads to, as long as those probabilities are kept positive, but only by the EAC on those classes, and 2) The EAC may depend on the cost structure at the transient states. In Section 4, we provide an example in connection with the question of the feasibility of a vanishing discount approach for exponential risk-sensitive models; see [1] and references therein for the vanishing discount approach in the risk-null model. Jaquette [15] proved that the optimal policies for the control problem with exponential discounted cost and finite state space are not stationary, in general. He showed, however, that for the mentioned control problem, there must exist optimal policies that are ultimately stationary, i.e., that apply the same decision function after a determined stage. Our example consists of a CMC with infinite state space for which the EDC control problem does not have ultimately stationary optimal policies, thus showing that for infinite state space models, the mentioned control problem is in gen-
eral even "more non-stationary" than it is for finite models. Moreover, the CMC in our example satisfies a simultaneous Doeblin condition, an assumption that enables the vanishing discount approach in the risk-null context. Therefore, the mentioned example strongly suggests that applying a vanishing discount approach in the risk-sensitive models might be even harder than what Jaquette's results had already indicated; see [7] for some recent results in that direction. Finally, we end this paper with an appendix including some auxiliary technical results.

2 Description of the Model.

Consider a stochastic dynamic system, specified by $(X, A, \{A(i) : i \in X\}, P, c)$, where:

a) $X$, the state space, is a discrete set. We will take $X = \{1, 2, \ldots, N\}$ when $X$ is finite and $X = \{1, 2, \ldots\}$ in the infinite case.

b) $A$, the action or control space, is a finite set.

c) $A(i)$, the set of admissible actions at $i$, is a subset of $A$. The set of admissible state-action pairs is defined as $K := \{(i, a) : i \in X, a \in A(i)\}$.

d) $P = \{P(\cdot | i, a) : (i, a) \in K\}$, is a transition probability on $K \times 2^X$, where $2^X$ is the family of all subsets of $X$. For brevity, sometimes we will write $P(j | i, a)$ instead of $P(\{j\} | x, a)$.

e) $c : K \to \mathbb{R}$ is the one-stage cost function. We will assume that $c(\cdot, \cdot) \geq 0$ and that $K := \{c(i, a) : (i, a) \in K\} < \infty$.

The above defined CMC represents a stochastic dynamical system observed at times (or events) $t \in \mathbb{N} := \{0, 1, 2, \ldots\}$. The evolution of the system is as follows. Let $X_t$ denote the state at time $t \in \mathbb{N}$, and $A_t$ the action chosen at that time. If at decision epoch $t$ the system is in state $X_t = i \in X$, and the control $A_t = a \in A(i)$ is chosen, then (i) a cost $c(i, a)$ is incurred, and (ii) the system moves to a new state $X_{t+1}$ according to the probability distribution $P(\cdot | i, a)$. Once the transition into the new state has occurred, a new action is chosen, and the process is repeated. We will take the stochastic processes $(X_t)$ and $(A_t)$ as given by the coordinate functions defined on $(X \times A)^\infty$ in the usual way; for more details, see [1, 12, 19]. For simplicity we will often denote $C_t := c((X_t, A_t))$.

Let $\Pi$ denote the set of all admissible (possibly randomized and history dependent) policies (see [1]), and $\mathcal{F}$ the set of admissible decision functions, i.e., functions from $X$ to $A$ such that $f(i) \in A(i)$ $\forall i \in X$. We will distinguish two subclasses of policies: the class $\Pi_{MD}$ of Markovian deterministic policies and the class $\Pi_{SD}$ of stationary deterministic policies [1]. The stationary deterministic policy determined by $f \in \mathcal{F}$ will be denoted by $f^\infty$. For each policy $\pi \in \Pi$ and each initial state $i \in X$ we define in the usual way a probability measure $P^\pi_t$ on $\Omega := (X \times A)^\infty$ and denote the corresponding expectation operator by $E^\pi_t$ [1, 17, 19]. When $\pi = f^\infty$ we simply write $P^\pi_t$ and $E^\pi_t$ respectively.

The (risk-null) average cost due to a policy $\pi \in \Pi$ and initial state $i \in X$ will be denoted by $\phi^\pi(i) := \limsup_{n \to \infty} \frac{1}{n} \left( E^\pi_t \left[ \sum_{t=0}^{n-1} C_t \right] \right)$.

The certainty equivalent of a random variable $Z$ with respect to the disutility function $U_\gamma$, is defined as

$$E(\gamma, Z) := \frac{1}{\gamma} \log \left( E[e^{\gamma Z}] \right).$$

Heuristically, a decision maker with utility function $U_\gamma$ is indifferent between the random (thus uncertain) cost $Z$ and the (certain) cost $E(\gamma, Z)$.

3 The exponential average cost: continuity counterexamples

In this section we restrict our attention to the finite state space model. We study the EAC corresponding to a fixed stationary deterministic policy. First, a result is stated which provides a method for computing the EAC in the general non-irreducible case, via the classification of states into self-communicating classes. Then, we use that result to examine some examples which illustrate the effect of risk-sensitivity in the non-irreducible case.

Since we will study the EAC for a fixed stationary deterministic policy, we will consider a simplified model (a Markov cost/reward chain) whose elements are: the state space $X = \{1, \ldots, N\}$, a stochastic matrix $P = (P_{ij})_{i,j=1}^N$, and a cost vector $c = (c(1), \ldots, c(N))$ with non-negative components. The EAC for the Markov cost process is then defined as the long-run average of the certainty equivalents (of the partial costs):

$$J(\gamma, i) = \limsup_{n \to \infty} \frac{1}{n} \log \left( E_i \left[ \exp \left( \sum_{t=0}^{n-1} c(X_t) \right) \right] \right).$$

where $E_i$ is the expectation operator on $(X, 2^X)^\infty$, induced by the transition probability matrix $P$ and the initial state $i \in X$. Moreover, if we define the disutility matrix $\bar{P}(\gamma)$ by means of $\bar{P}(\gamma)(i, j) := (P(i, j)e^{\gamma c(i)})$ (sometimes we will write just $P$ instead of $\bar{P}(\gamma)$), then it is easy to check (see for example [2] or [3]) that the following alternative expression of the EAC holds:

$$J(\gamma, i) = \limsup_{n \to \infty} \frac{1}{n} \log \left( \sum_{j=1}^N \bar{P}^n(i, j) \right).$$

It is a well known fact that when the transition probability matrix $P$ is irreducible, the EAC does not depend
on the initial state. Moreover,
\[ J(\gamma, i) = \frac{1}{\gamma} \log \lambda =: J(\gamma), \]
for every \( i \in X \), where \( \lambda \) is the dominant eigenvalue of the irreducible matrix \( \tilde{P}(\gamma) \); see \([14, 9, 3]\). We will state a result which shows that a similar characterization holds as well for the case in which \( P \) is not irreducible (although in this case, as expected, the EAC does depend on the initial state.) First, we recall some basic notation from the theory of probability and nonnegative matrices. For \( i, j \in X \), we write \( i \rightarrow j \) and say that \( i \) leads to \( j \) (or \( j \) is accessible from \( i \)), when \( P^n(i, j) = P(X_n = j | X_0 = i) > 0 \) for some \( n \geq 1 \) (observe that with this definition it may happen that \( i \leftrightarrow i \), in which case we say that \( i \) is an irrelevant state.) Also, \( i \leftrightarrow j \), which is read \( "i \) and \( j \) communicate", means that \( i \rightarrow j \) and \( j \rightarrow i \). Similarly, for \( C \subset X \), \( i \rightarrow C \) ("the class \( C \) is accessible from \( i \)) means that \( i \rightarrow j \) for some \( j \in C \). A class of states \( C \subset X \) is called self-communicating if a) \( i \leftrightarrow j \) for every \( i, j \in C \), and b) there does not exist a subclass \( C' \subset X \) and containing \( C \) such that (a) holds for \( C' \). We will denote the spectral radius of a square matrix \( A \) by \( \rho(A) \).

We will assume that the transition probability matrix \( P \) is put in its irreducible canonical form (cf. \([13]\)):
\[ P = \begin{pmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & P_m & 0 \\ W_1 & W_2 & \cdots & W_m & T \end{pmatrix}. \]

In (2), the matrices \( P_r, r = 1, \ldots, m \) are the irreducible transition probability matrices corresponding to the closed and irreducible classes \( C_r, r = 1, \ldots, m \). The entries of the matrices \( W_r, r = 1, \ldots, m \), are the one step transition probabilities from transient to recurrent states. We also assume that \( T \), the matrix whose entries are the one-step transition probabilities among transient states, is in turn in the form
\[ T = \begin{pmatrix} P_{m+1} & 0 & 0 & \cdots & 0 \\ G_{21} & P_{m+2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ G_{i-1,m,1} & G_{i-1,m,2} & \cdots & P_{t-1} & 0 \\ G_{i,m,1} & G_{i,m,2} & \cdots & G_{i,m-l,m-1} & P_l \end{pmatrix}, \]
where each matrix \( P_r, r = m+1, \ldots, l \), is irreducible or equal to the one by one zero matrix. If \( P_r \neq 0 \) for some \( r = m+1, \ldots, l \), then the entries of \( P_r \neq 0 \) are the transition probabilities among states in a self-communicating class \( C_r \) of transient states. If \( P_r = (P_{ij}) = 0 \), we say that \( C_r = \{j\} \) is an irrelevant class.

Before stating the announced representation of the EAC, we will state two lemmas which provide additional insight into the behavior of the EAC. For a proof of the three following results and further details, see \([3]\).

**Lemma 1** For \( i, j \in X, \ i \rightarrow j \Rightarrow \text{sgn}(\gamma) J(\gamma, i) \leq \text{sgn}(\gamma) J(\gamma, j) \).

**Remark 1** (a) It follows from Lemma 1 that \( J(\gamma, i) = J(\gamma, j) \) when \( i \leftrightarrow j \), that is, the EAC is constant within a self-communicating class.

(b) Lemma 1 also hold when the state space \( X \) is infinite.

**Lemma 2** The EAC satisfies the equation
\[ J(\gamma, i) = \max \{\text{sgn}(\gamma) J(\gamma, j) : i \rightarrow j\}, \quad i \in X, \]

**Theorem 1** If the probability matrix \( P \) is expressed in the irreducible canonical form, then for each \( i \in X \) we have
\[ J(\gamma, i) = \frac{1}{\gamma} \max \{\log \rho(P_r) : i \rightarrow C_r\} \]

**Remark 2** Two remarkable differences between the risk-sensitive and the risk-neutral model become apparent from Theorem 2:
(1) The EAC when beginning at a transient state is not a typical average of the EAC's over the closed classes that are accessible from the initial state; and
(2) The EAC may depend on the cost structure at the transient states.

**Remark 3** Notice that the proof of the theorem is still valid if we substitute \( \limsup \) by \( \liminf \) wherever the first one appears. We deduce from that observation that, for the finite state space model, the limit exists in the definition of the EAC.

It can be proved that if we define the EAC for \( \gamma = 0 \) as the risk-neutral average cost, i.e., \( J(0, i) := \phi \forall i \in X \), then \( J(\cdot, i) \) turns out to be continuous at \( \gamma = 0 \) for every \( i \in X \) (cf. \([2, 3]\)). Our first example demonstrates how characteristic (1) in Remark 2 may cause the mentioned property of the EAC not to hold in the non-irreducible case. The example considers a model with a transient state \( x \) which leads to more than one closed class.

**Example 1.** Consider the cost process with state space \( X = \{1, 2, 3\} \), transition probability matrix
\[ P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p & 1-p & 0 \end{pmatrix}, \]
and cost vector \( c = (1,3,5) \). From Theorem 1 we see that
\[
J(\gamma, 3) = \begin{cases} 
\max\{c(1), c(2)\} = 3 & \text{if } \gamma > 0, \\
\min\{c(1), c(2)\} = 1 & \text{if } \gamma < 0,
\end{cases}
\]
so that \( \lim_{\gamma \to 0^+} J(\gamma, 3) = 3 \neq 1 = \lim_{\gamma \to 0^-} J(\gamma, 3) \).

On the other hand, it is easy to check that \( \phi(3) = p c(1) + (1-p) c(2) = 3 - 2p \). Thus, \( J(\cdot, 3) \) is neither left nor right continuous at zero.

**Remark 4** Note that \( J(\cdot, 3) \) does not depend on the probabilities \( p \) and \( 1-p \) of going from the transient state 3 to the closed classes \{1\} and \{2\} respectively, as long as both probabilities are positive.

It is a well known fact that if a Markov cost chain has a unichain structure, i.e. only one irreducible closed class and possibly some transient states, then the standard (risk-neutral) average cost does not depend on the initial state. The following example shows how feature (2) in Remark (2) determines the mentioned property not to be true in general for the EAC, when the risk-sensitivity coefficient is large enough.

**Example 2.** Consider the cost process with state space \( X = \{1,2\} \), transition matrix
\[
P = \begin{pmatrix}
1 & 0 \\
1-p & p
\end{pmatrix},
\]
and cost vector \( c = (c(1), c(2)) \) such that \( c(1) < c(2) \). From Theorem 2 we have that 1) \( J(\gamma, 1) = c(1) \) for every \( \gamma \), and 2) \( J(\gamma, 2) = \max\{c(1), \frac{1}{\gamma} \log p(e^{\gamma c(2)})\} \) for \( \gamma > 0 \), and \( J(\gamma, 2) = \min\{c(1), \frac{1}{\gamma} \log p(e^{\gamma c(2)})\} \) for \( \gamma < 0 \), that is
\[
J(\gamma, 2) = \begin{cases} 
c(1) & \text{if } \gamma < \frac{\log p}{c(1) - c(2)}, \\
\frac{1}{\gamma} \log p & \text{if } \gamma \geq \frac{\log p}{c(1) - c(2)}.
\end{cases}
\]
Thus, if \( \gamma > \frac{\log p}{c(1) - c(2)} \), then the EAC does depend on the initial state. On the other hand, we can observe that the EAC is better behaved for small values of \( \gamma \):
\( J(\gamma, 1) = J(\gamma, 2) \) for \( \gamma \leq \frac{\log p}{c(1) - c(2)} \).

The following corollary to Theorem 2 shows that, as expected, the cost structure at the transient states does not influence the EAC when \( \gamma \) is close to zero.

**Corollary 1** Assume that \( P \) is expressed in its irreducible canonical form (as in Theorem 1). Then there exists \( \gamma_0 > 0 \) such that for every transient state \( i \) and \( \gamma \in (-\gamma_0, \gamma_0) \),
\[
J(\gamma, i) = \frac{1}{\gamma} \max \{ \log \rho(\bar{P}_k) : i \to C_r \}
\]
and \( C_r \) is ergodic.

**Proof:** Since the proofs of the corollary for \( \gamma > 0 \) and \( \gamma < 0 \) are very similar, we carry out only the former one. Take any of the blocks \( \bar{P}_k \) in the irreducible canonical form of \( T \), corresponding to a class of transient states, that is, with \( m+1 \leq s \leq t \). Since \( P_k \) is strictly substochastic, \( \rho(P_k) < 1 \), and consequently \( \log \rho(P_k) < 0 \). Now, \( \bar{P}_k e^{-\gamma K} P_k \) implies that \( \rho(\bar{P}_k) e^{-\gamma K} \rho(P_k) \), that is,
\[
\frac{1}{\gamma} \log(\bar{P}_k) \leq K + \frac{1}{\gamma} \log(\rho(P_k)).
\]

Therefore,
\[
\lim_{\gamma \to 0^+} \frac{1}{\gamma} \log \rho(\bar{P}_k) \leq K + \lim_{\gamma \to 0^-} \frac{1}{\gamma} \log \rho(P_k) = -\infty,
\]
and the claim follows from Theorem 1 and the fact that there is only a finite number of transient states.

**Remark 5** It follows from the corollary that, under a unichain structure assumption, the EAC does not depend on the initial state for \( |\gamma| \) small enough. Furthermore, similarly as in the irreducible case, the EAC converges to the risk-neutral average cost when \( \gamma \) goes to zero.

7. The exponential discounted cost: an example of a non-ultimately stationary optimal policy.
Motivated by well established results in the risk-neutral literature, it is natural to ask whether a vanishing discount approach is possible or not in the risk-sensitive case: the idea is to obtain the EAC from the exponential (discounted cost)
\[
V^\gamma_i := E^\gamma \left[ \exp(\sum_{t=0}^{\infty} e^r C_t) \right],
\]
where \( \beta \in (0,1) \); see [1] for the risk-null case. Jaquette [15] showed that the optimal policies of the control problem corresponding to the performance function \( V^\gamma_i \) are not stationary in general. More precisely, he gave an example of an irreducible CMC with finite state space for which the optimal policy is not stationary. On the other hand, stationary optimal policies can be obtained from the (exponential) average cost optimality equation for that kind of models, (cf.[2, 3]. The previously mentioned facts strongly suggest that the conditions under which a vanishing discount approach is valid for exponential criteria, must be more restrictive than those in the risk-null model, see [7]. In this section, we present an example that further supports that belief, as we explain right now. In Jaquette’s example, although not stationary, the optimal policy –with respect to (4)– is ultimately stationary, i.e., it is of the form \( \pi = \{f_1, f_2, \ldots\} \) with \( f_k = f_{k+1} = \cdots \) for some \( k \in \mathbb{N} \). In fact, Jaquette proves that for any CMC with finite state space the optimal policy must be ultimately stationary. Our example shows that assertion not to be
true for CMC's with infinite state space. Moreover, the CMC in the example satisfies a simultaneous Doeblin condition, an assumption commonly used to apply the vanishing discount scheme in the risk-null framework. Thus, the exponential discounted optimal control problem for the infinite state space model is even “more non-stationary” than that with finite state space. As a consequence, a vanishing discount method might be harder to obtain in the infinite case. For the rest of this section, we will restrict our attention to the risk-averse case, i.e., we will assume \( \gamma > 0 \).

Let \( \preceq \) denote the stochastic order induced by the (dis)utility function \( U_\gamma \), that is,

\[
X \preceq Y \iff E[e^{\gamma X}] \leq E[e^{\gamma Y}]
\]

for \( X \) and \( Y \) in the space of real bounded random variables. It is not hard to see that the order \( \preceq \) is not homogeneous, i.e., for some \( \beta > 0 \) and random variables \( X, Y \), we may have \( X \preceq Y \) and \( \beta X \succ \beta Y \) (where \( \succ \) means “not \( \preceq \).”) The key idea in the design of the example is to use the non-homogeneity of \( \preceq \) to define real random variables \( Z, W_t, t = 2, 3, \ldots \) and a sequence of integers \( \{t_i : i = 2, 3, \ldots \} \subset \{2, 3, \ldots \} \) such that \( t_i \uparrow \infty \) and

\[
E[e^{\beta t Z}] \geq (\prec) E[e^{\beta t Y}] \quad \text{for} \quad t \leq (\succ) t_i. \tag{5}
\]

The first step is to take independent real random variables \( Z \) and \( Y_n, n = 1, 2, \ldots \) with distributions given by

\[
Z \sim \frac{2}{3}(0) + \frac{1}{3}(3) \quad \text{and} \quad Y_n \sim \frac{1}{2}(0) + \frac{1}{2}(2 + \frac{1}{n}).
\]

Now, for each \( n \in \mathbb{N} \) define the function \( G_n : [0, \infty) \rightarrow \mathbb{R} \) by

\[
G_n(\lambda) := E[e^{\lambda Y_n}] - E[e^{\lambda Z}].
\]

We prove in Proposition (3) in the Appendix, that there exists a sequence of positive numbers \( \lambda_n \) decreasing to zero, such that

\[
G_n(\lambda) > (\prec) = 0 \iff \lambda < (\succ) = \lambda_n.
\]

Thus, if we define the subsequence \( \{\gamma_k\} := \{\lambda_n\} \) recursively by \( \gamma_1 = \max(\lambda_n : \lambda_n < \beta) \) and \( \gamma_k = \max(\lambda_n : \lambda_n < \beta^{k-1}, \lambda_n < \gamma_{k-1}) \) for \( k \geq 2 \), then, it is easy to verify that \( Z \) and the sequences \( W_t := Y_{t+n}, t_i = \max\{t : \beta^t \geq \gamma_i\} \) satisfy (5). We are now prepared to construct the announced example.

Example 4. Consider the CMC defined by

\[
X := \mathbb{N}, \Delta := \{a_1, a_2\}, A(1) = \{a_1\}, \quad \text{and} \quad A(i) = \{a_i\},
\]

\[
P(1 \mid 1, a_1) = P(1 \mid i, a_1) = 1,
\]

\[
P(i \mid i, a_2) = P(1 \mid i, a_2) = \frac{1}{2},
\]

\[
c(i, a_1) \sim Z, \quad c(1, a_1) \sim Z \quad \text{and} \quad c(i, a_2) \sim W_i,
\]

for \( i \geq 2 \), where \( Z \) and \( W_i, i = 1, 2, \ldots \) are the random variables defined above, and the random variables \( c(1, a_1), c(i, a_1), \) and \( c(i, a_2), i = 1, 2, \ldots \), are independent.

Now, set \( \gamma = 1 \). By the way in which the random variables \( Z \) and \( W_i \) were defined, it is clear that

\[
E[e^{\beta(1, a_2)}] = E[e^{\beta W_i}], \quad E[e^{\beta^2(1, a_1)}] = E[e^{\beta Z}],
\]

and therefore

\[
E[e^{\beta^t c(i, a_2)}] \leq (\succ) E[e^{\beta^t c(i, a_1)}]
\]

for \( t \leq (\succ) t_i \). Since the random variables \( c(1, a_1), c(i, a_1), \) and \( c(i, a_2), i = 1, 2, \ldots \), are independent, the non-ultimately stationary policy \( \pi^* = \{f_t : t = 1, 2, \ldots \} \) given by

\[
f_t(i) = \begin{cases} a_1 & \text{if } t > t_i, \\ a_2 & \text{if } t \leq t_i, \end{cases}
\]

is optimal. Moreover, no ultimately stationary policy can be optimal, since \( t_i \uparrow \infty \) as \( i \uparrow \infty \).

Remark 6 Note that the optimal policy \( \pi^* \) is ultimately stationary for each initial state \( i \) fixed. Using the same method as above, a similar (but more technically involved) example can be constructed where ultimately stationary optimal policies do not exist even for a fixed initial state.

References


Appendix

The proof of the following proposition can be found in ([13]).

**Proposition 1** If a non-negative square matrix \( A \) is in the block form

\[
A = \begin{pmatrix}
A_{11} & 0 & 0 & \cdots & 0 \\
A_{21} & A_{22} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
A_{m-1,1} & A_{m-1,2} & \cdots & A_{m-1,m-1} & 0 \\
A_{m1} & A_{m2} & \cdots & A_{m,m-1} & A_{mm}
\end{pmatrix},
\]

where the diagonal blocks \( A_{ii} \) are square matrices, then

\[
\rho(A) = \max \{\rho(A_{ii}) : 1 \leq i \leq m\} \quad (6)
\]

The following proposition follows from [8, Lema 1.2.15]

**Proposition 2** Assume \( X \) is finite and let \( \mu \) be an arbitrary measure on \((X, 2^X)\). If \( Z \) and \( Z_n, n = 1, 2, \ldots \) are real random variables defined on \( X \) such that

\[
\limsup_{n \to \infty} Z_n = Z,
\]

then

\[
\limsup \left( \int Z_n d\mu \right)^{\frac{1}{n}} = \text{ess sup} (Z).
\]

**Remark 1** It can be proved that the proposition also holds with \( \limsup \) replaced by \( \liminf \) everywhere, see [3]. Consequently, the proposition is true as well if we take \( \lim \) instead of \( \limsup \) everywhere.

We provide the proof of the following proposition in [3].

**Proposition 3** For each \( n \in \mathbb{N} \) define the function \( G_n : [0, \infty) \to \mathbb{R} \) by

\[
G_n(\lambda) = \left( \frac{1}{2} + \frac{1}{2} e^{(2+\frac{1}{n})\lambda} \right) - \left( \frac{1}{3} + \frac{1}{3} e^{3\lambda} \right)
\]

\[
= e^{3\lambda} \left( \frac{1}{2} e^{-(1+\frac{1}{n})\lambda} - \frac{1}{3} \right) - \frac{1}{6}.
\]

Then, there exists a sequence of positive numbers \( \{\lambda_n : n \geq 2\} \) decreasing to 0 such that

\[
G_n(\lambda) \begin{cases} 
= 0 & \text{if } \lambda = \lambda_n, \\
> 0 & \text{if } \lambda < \lambda_n, \\
< 0 & \text{if } \lambda > \lambda_n,
\end{cases} \quad (7)
\]

for \( n \geq 2 \).