Proceedings of the 36th
Conference on Decision & Control
San Diego, California USA • December 1997

CONTROLLED MARKOV CHAINS WITH RISK-SENSITIVE EXPONENTIAL AVERAGE COST CRITERION[1]

Agustin Brau[2]
Department of Mathematics, The University of Arizona
and
Emmanuel Fernández-Gaucherand
Systems and Industrial Engineering Department, The University of Arizona

1 Introduction

In this paper we are concerned with the exponential risk-sensitive version of the standard average cost criterion for controlled Markov chains (CMC's); see also [3], [11], [5]. Our presentation is more mathematically rigorous than that in [7], and our proof techniques are more self-contained and perhaps more intuitive than those in [3]. Furthermore, we extend some results in [3], [7] to the countable state space case. In addition, we consider optimization within the general set of randomized policies, and not only within the restricted class of Markovian deterministic policies. We model risk sensitivity as being given by an *exponential disutility* function $\mathcal{U}_\gamma(x) = (\text{sgn}\gamma)e^{\gamma x}$, where γ is the constant risk-sensitivity coefficient, see [13], [7]. First, in Section 2 we give the basic definitions and notation of our model. Section 3 presents and briefly analyzes alternative definitions of the EAC. Section 4 is devoted to a more detailed discussion of Howard and Matheson's definition of EAC. Finally, in Section 4 we show that, similarly to the risk-neutral case, the optimal EAC satisfies an *optimality equation*; see also [3], [11].

2 The Model

Let us consider a discrete time, stationary, controlled Markov chain (CMC) especified by the tuple $(\mathbf{X}, \mathbf{A}, \mathbf{P}, c)$, where

a) $\mathbf{X}$, the state space, is a *countable* set, say $\mathbf{X} = \{1, 2, \dots\}$. The elements of $\mathbf{X}$ are called states.

b) $\mathbf{A}$, the action (or control) set, is a finite set. The set of admissible state-action pairs is defined as $\mathbf{K} := X \times \mathbf{A}$.

c) $\mathbf{P}$, the transition kernel, is a family of transition probabilities on $\mathbf{X}$ given $\mathbf{K}$:

$$\mathbf{P} = \{P(\cdot \mid x, a) : (x, a) \in \mathbf{K}\}.$$

We will also denote $p_{xx'}(a) := P(x' \mid x, a)$.

d) $c : \mathbf{K} \longrightarrow \mathbb{R}$ is the one-stage cost function. We will assume that c is non-negative and bounded: $0 \leqslant c(x, a) \leqslant K$ for some constant $K \in (0, \infty)$

The above defined CMC represents a stochastic dynamical system observed at times $t = 0, 1, 2, \dots$. The evolution of the system is as follows. Let X_t denote the state at time $t \in \mathbb{N}_0$, and A_t the action chosen at that time. If at decision epoch t the system is in state $X_t = x \in \mathbf{X}$, and the control $A_t = a \in \mathbf{A}$ is chosen, then (i) a cost $c(x, a)$ is incurred, and (ii) the system moves to a new state X_{t+1} according to the probability distribution $P(\cdot \mid x, a)$. Once the transition into the new state has ocurred, a new action is chosen, and the process is repeated. For more details, see [1], [6], and [14].

We will use the following standard notation: $\mathbf{\Pi}$ and $\mathbf{\Pi}_{SD}$, respectively, denote the sets of (randomized) admissible and stationary deterministic policies. The (risk-null) average cost function due to a policy $\pi \in \mathbf{\Pi}$ will be denoted by

$$\phi^\pi(i) := \limsup_{n\to\infty} \frac{1}{n} E_i^\pi \left[\sum_{t=0}^{n-1} C_t\right] \tag{1}$$

for $i \in \mathbf{X}$, where we denote for brevity $C_t := c((X_t, A_t))$.

3 Defining an Exponential Risk-sensitive (ERS) Average Cost Criterion

For an Expected utility criterion, the value function Φ_π corresponding to a policy π is given by the expected value of a certain measurable functional defined

[1]Research partially supported by the National Science Foundation under grant NSF-INT 9602939

[2]The work of this author was partially supported by Universidad de Sonora and CONACYT, México.

on the space of sample paths: $\Phi_\pi(i) = E_i^\pi[F]$, $i \in \mathbf{X}$, where $F : \Omega \to [-\infty, \infty]$, and $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$. For that type of criterion, the corresponding ERS version can be defined as $E_i^\pi[\mathcal{U}_\gamma(F)]$ or $\mathcal{U}_\gamma^{-1} E_i^\pi[\mathcal{U}_\gamma(F)]$, which are respectively the expected disutility and the certain equivalent of F with respect to $\mathcal{U}_\gamma$. Jaquette [8, 9] analized the ERS criterion corresponding to the discounted cost, $F = \sum_{t=0}^\infty \beta^t C_t$, and Denardo and Rothblum ([2]) studied that corresponding to the total cost, $F = \sum_{t=0}^\infty C_t$.

Clearly, *the average cost ϕ^π is not an expected utility criterion.* Thus, the point arises here as how to define an ERS version of ϕ^π. Now, take into account that the numerical sequence $\frac{1}{n} E_i^\pi\left[\sum_{t=0}^{n-1} C_t\right]$ in (1) may also be written as $E_i^\pi\left[\frac{1}{n}\sum_{t=0}^{n-1} C_t\right]$, and consider the disutility function $\mathcal{U}_\gamma$ instead of the linear disutility $\mathcal{U}_0(x) \equiv x$. Then, we can initially think of the following three index functions as candidates for the ERS version of the standard average cost (1):

$$\bar{J}^\pi(i,\gamma) := \limsup_{n\to\infty} \frac{1}{n} E_i^\pi\left[\mathcal{U}_\gamma\left(\sum_{t=0}^{n-1} C_t\right)\right] = \limsup_{n\to\infty} \frac{1}{n} E_i^\pi\left[(\operatorname{sgn}\gamma)\exp\left(\gamma\sum_{t=0}^{n-1} C_t\right)\right], \tag{2}$$

$$\tilde{J}^\pi(i,\gamma) := \limsup_{n\to\infty} \mathcal{U}_\gamma^{-1}\left(E_i^\pi\left[\mathcal{U}_\gamma\left(\frac{1}{n}\sum_{t=0}^{n-1} C_t\right)\right]\right) = \limsup_{n\to\infty} \frac{1}{\gamma}\log\left\{E_i^\pi\left[\exp\left(\gamma\frac{\gamma}{n}\sum_{t=0}^{n-1} C_t\right)\right]\right\}. \tag{3}$$

$$J^\pi(i,\gamma) := \limsup_{n\to\infty} \frac{1}{n}\mathcal{U}_\gamma^{-1}\left\{E_i^\pi\left[\mathcal{U}_\gamma\left(\sum_{t=0}^{n-1} C_t\right)\right]\right\} = \limsup_{n\to\infty} \frac{1}{n}\frac{1}{\gamma}\log\left\{E_i^\pi\left[\exp\left(\gamma\sum_{t=0}^{n-1} C_t\right)\right]\right\}, \tag{4}$$

Remark 1 *Note that in the first two above functions we first introduce risk-sensitivity (by considering the expected disutility and the certain equivalent of the total cost up to time n in (2) and (4) respectively) and then we take the average. Conversely, in (3) we first take the average and then the certain equivalent. Observe also that $\tilde{J}^\pi$ and the function obtained from it considering expected disutilities instead of certain equivalents induce the same order relation among policies, because $\mathcal{U}_\gamma^{-1}$ is increasing. Thus, we disregard that possibility at the outset. Next, we analyze the strengths and defficiencies of these alternative definitions.*

3.1 Candidate 1: $\bar{J}^\pi$.

First we will see that, at least in the interesting situations, (2) is a trivial function. Let's consider the case $\gamma > 0$. Let $\pi \in \mathbf{\Pi}$ and assume that $\phi^\pi(i) = \alpha > 0$. Then there exists a sequence of positive integers $\{n_k\} \uparrow \infty$ such that $\frac{1}{n_k}\sum_{t=0}^{n_k-1} E_i^\pi[C_t] > \frac{1}{2}\alpha$. Thus, we have

$$\frac{1}{n_k}\exp\left\{\gamma\sum_{t=0}^{n_k-1} E_i^\pi[C_t]\right\} > \frac{1}{n_k}\exp\left(\gamma\frac{\alpha}{2}n_k\right) \underset{k\to\infty}{\longrightarrow} \infty,$$

which implies that

$$\limsup_{n\to\infty}\frac{1}{n}\exp\left\{\gamma E_i^\pi\left[\sum_{t=0}^{n-1} C_t\right]\right\} = \infty.$$

Now, by Jensen's inequality,

$$\frac{1}{n}E_i^\pi\left[\exp\left\{\gamma\sum_{t=0}^{n-1} C_t\right\}\right] \geqslant \frac{1}{n}\exp\left\{\gamma E_i^\pi\left[\sum_{t=0}^{n-1} C_t\right]\right\},$$

and we conclude that $\bar{J}^\pi(i,\gamma) = \infty$. If $\gamma < 0$, we can show similarly that for every policy π such that $\phi^\pi(i) = \alpha > 0$, we have $\bar{J}^\pi(i,\gamma) = 0$. Thus, $\pi \to \bar{J}^\pi$ does not discriminate among policies.

Remark 2 *The assumption $\phi^\pi(i) = \alpha > 0$ is not restrictive for the average cost criterion. If for some policy π it holds that $\phi^\pi(i) = \alpha = 0$, then $E_i^\pi[\sum_{t=0}^\infty C_t]] < \infty$. In that case, a total cost criterion is clearly more adequate for the decision problem.*

3.2 Candidate 2: $\tilde{J}^\pi$.

By its form, $\tilde{J}^\pi$ seems to capture the essence of both the average cost as a long-run behaviour performance index and risk-sensitivity as sensitivity to randomness. To gain more insight into the nature of (3), consider the case of a policy $\pi \in \mathbf{\Pi}$ and an initial state $i \in \mathbf{X}$ for which $\lim \frac{1}{n}\sum_{t=0}^{n-1} C_t$ exists and is equal to a random variable D (P_i^π-a.e.). Then, applying the dominated convergence theorem, we see that $\tilde{J}^\pi(i,\gamma)$ is just the certain equivalent of D with respect to $\mathcal{U}_\gamma$. In particular, when the state space is finite and i is a recurrent state under a stationary deterministic policy $\pi = \{f, f, \dots\}$, D is (almost surely) constant and $\tilde{J}^\pi(i,\gamma) = \phi^\pi(i,\gamma)$ $\forall\gamma$. The reduction of the risk-sensitive index to the risk-neutral one in the previous ergodic situation is intuitevely consistent with the fact that randomness "disappears" in the long-run. That appealing property of $\tilde{J}^\pi$ might turn out to be also its weakness, because ergodicity is, in general, the condition that allows the study of the long-run behaviour of stochastic processes. Thus, it appears that this candidate holds little promise.

3.3 Candidate 3: J^π.

This index function was introduced by Howard and Matheson [7] in the particular context of a finite state-action unichain model, and it has been widely accepted in the literature as the ERS average cost [3], [11]. Recently, it has been also studied in connection with robust control theory [3], [11], [5]. A very appealing consequence for this candidate is that useful optimality equations can be obtained. In the next section we will discuss some of the main properties and consequences of using (4).

4 Howard-Matheson's exponential average cost criterion

Throughout this section we will make the two following additional assumptions:

Assumption 1: The state space is finite:$\mathbf{X} = \{1, \dots, N\}$

Assumption 2: For every $(f^\infty) \in \mathbf{\Pi}_{SD}$, the corresponding probability transition matrix $P_f = (p_{ij}(f(i)))$ is primitive.

For $(f^\infty) \in \mathbf{\Pi}_{SD}$ let us denote by Q_f the matrix with entries $q_{ij}(f) := e^{\gamma c(i,f(i))} p_{ij}(f(i))$. For the sake of brevity let us also denote the exponential disutility and certain equivalent of the total cost up to time n respectively by $U_n^f(i,\gamma)$ and $J_n^f(i,\gamma)$, that is

$$U_n^f(i,\gamma) := E_i^f \left[\mathcal{U}_\gamma \left(\sum_{t=0}^{n-1} C_t \right) \right],$$

and

$$J_n^f(i,\gamma) := \mathcal{U}_\gamma^{-1} \left(U_n^f(i,\gamma) \right).$$

Finally, when $\pi = f^\infty$, we write J^f instead of J^π, and hence we may write definition (4) as $J^f := \limsup_{n\to\infty} \frac{1}{n} J_n^f(i,\gamma)$.

Under the previous assumptions, Howard and Matheson [7] showed that the sequence $J_n^f(i,\gamma)$ actually converges, is independent of the initial state i, and

$$\begin{aligned} J^f(i,\gamma) &= \lim_{n\to\infty} \frac{1}{n} J_n^f(i,\gamma) \\ &= \frac{1}{\gamma} \log \lambda =: J^f(\gamma) \end{aligned} \tag{5}$$

where λ is the dominant eigenvalue of the primitive positive matrix Q_f. The above limiting relation was also established by Miller [12] to show (in terms of our problem) that the sequence $\frac{1}{n}\sum_{t=0}^{n-1} c(X_t, f(X_t))$ obeys a large deviations principle.

Howard and Matheson also showed that if $w = (w(1), \dots, w(N))$ is the unique eigenvector corresponding to λ such that $w(N) = (\mathrm{sgn}\gamma)$, then

$$\lim_{n\to\infty} \left[J_n^f(i,\gamma) - J_n^f(N,\gamma) \right] = H(i,\gamma)$$

$\forall i \in \mathbf{X}$, where $H(i,\gamma) := \mathcal{U}_\gamma^{-1}(w(i))$. The above equation says that the function $H(\cdot,\gamma)$ gives the *relative certain equivalents with respect to the state* N. Moreover, since w is an eigenvector of Q_f corresponding to λ, we have that

$$\lambda w(i) = \sum_{j=1}^N q_{ij}(f) w(j)$$

$\forall i \in \mathbf{X}$, or equivalently, on taking account of (5) and the definition of H:

$$e^{\gamma J^f(\gamma)} = \sum_{j=1}^N q_{ij}(f) e^{\gamma H(i,\gamma)} \tag{6}$$

There is a clear resemblance between the above equation and the value equation for the risk-neutral average cost, under the present assumptions,

$$\phi^f + h(i) = c(i, f(i)) + \sum_{j=1}^N p_{ij}(f(i)) h(j), \tag{7}$$

where h is the corresponding relative value function, (see for example [1].) In the proposition below we show that the relation between equations (6) and (7) is in fact closer.

Since we will consider just one fixed policy $(f^\infty) \in \mathbf{\Pi}_{SD}$, for the sake of brevity we will remove the explicit dependance on f in every term. Accordingly, we will denote $P_{ij}(f(i)) = p_{ij}$, $c(i, f(i)) = c(i)$, $J^f(i,\gamma) = J(\gamma)$. On the other hand, we will introduce the explicit dependence on the risk-sensitivity coefficient γ of the eigenvalue λ and the eigenvector w of equation (6).

Proposition 1 *Under the Assumptions 1-2 we have*

$$\lim_{\gamma\to 0} J(\gamma) = \phi \quad \text{and} \quad \lim_{\gamma\to 0} H(\gamma, i) = h(i) \tag{8}$$

for every $i \in \mathbf{X}$, where h is the relative value function for the risk null case such that $h(N) = 0$.

Proof: : By denoting the eigenvector corresponding to $\lambda(\gamma)$ by $w(\gamma) = (w_1(\gamma), \dots, w_N(\gamma))$, and $u_i(\gamma) := (\mathrm{sgn}\gamma) w_\gamma(i)$, we obtain from that equation, after multiplying both sides by $\mathrm{sgn}\gamma$:

$$\sum_{j=1}^N p_{ij}(f) u_j(\gamma) e^{\gamma c(i)} = \lambda(\gamma) u_i(\gamma), \tag{9}$$

$i \in \mathbf{X}$. Now, u_i and λ are continuously differentiable (in fact analytic) functions of γ in a neighborhood of $\gamma = 0$ (see Kato [10, Ch.II].) Differentiating both sides of the equation above with respect to γ yields

$$\begin{aligned} &\sum_{j=1}^N p_{ij}(f) [u_j'(\gamma) + u_j(\gamma) c(i)] e^{\gamma c(i)}, \\ &\qquad = \lambda'(\gamma) u_i(\gamma) + \lambda(\gamma) u_i'(\gamma), \end{aligned}$$

and letting $\gamma \to 0$:

$$\sum_{j=1}^{N} p_{ij}(f)[\lim_{\gamma\to 0} u_j'(\gamma) + c(i) \lim_{\gamma\to 0} u_j(\gamma)] = \lim_{\gamma\to 0} \lambda'(\gamma) \lim_{\gamma\to 0} u_i(\gamma) + \lim_{\gamma\to 0} \lambda(\gamma) \lim_{\gamma\to 0} u'(\gamma). \quad (10)$$

Now, by letting $\gamma \to 0$ in the inequalities

$$\min_{i\in S} \sum_{j=1}^{N} p_{ij}(f) e^{\gamma c(i)} \leqslant \lambda \leqslant \max_{i\in S} \sum_{j=1}^{N} p_{ij}(f) e^{\gamma c(i)}$$

(see [4, Vol. II, Ch. XIII]), we obtain

$$\lim_{\gamma\to 0} \lambda(\gamma) = 1. \quad (11)$$

Moreover, also by taking the limit as $\gamma \to 0$ in both sides in (9) yields

$$P_f[\lim_{\gamma\to 0} u(\gamma)] = \lim_{\gamma\to 0} u(\gamma),$$

where $u(\gamma) = (u_1(\gamma), \dots, u_N(\gamma))$. Thus, the vector $\nu := \lim_{\gamma\to 0} u(\gamma)$ is a right eigenvector of P_f with respect to the eigenvalue 1, that is, an invariant vector. Furthermore, since $\nu_N = 1$, we must have that $\nu_i = 1$ for every $i \in S$, that is

$$\lim_{\gamma\to 0} u_i(\gamma) = 1 \quad (12)$$

By substituting (12) and (11) in (10), we get

$$\sum_{j=1}^{N} p_{ij}(f) \left[\lim_{\gamma\to 0} u_j'(\gamma) + c(i)\right] = \lim_{\gamma\to 0} \lambda'(\gamma) + \lim_{\gamma\to 0} u_i'(\gamma). \quad (13)$$

Observe that $u_N(\gamma) \equiv 1$ implies $\lim_{\gamma\to 0} u_N'(\gamma) = 0$.

Since the solution (ϕ, h) with $h(N) = 0$ of the value equation (7) is unique, we deduce from (13) that

$$\phi = \lim_{\gamma\to 0} \lambda'(\gamma) \text{ and } h(i) = \lim_{\gamma\to 0} u_i'(\gamma).$$

Finally, the proposition follows by observing that

$$\lim_{\gamma\to 0} J(\gamma) = \lim_{\gamma\to 0} \frac{1}{\gamma} \log \lambda(\gamma) = \lim_{\gamma\to 0} \frac{\lambda'(\gamma)}{\lambda(\gamma)} = \lim_{\gamma\to 0} \lambda'(\gamma),$$

and

$$\lim_{\gamma\to 0} H(\gamma, i) = \lim_{\gamma\to 0} \frac{1}{\gamma} \log(\text{sgn}\gamma) w_i(\gamma) = \lim_{\gamma\to 0} \frac{1}{\gamma} \log u_i(\gamma) = \lim_{\gamma\to 0} \frac{u_i'(\gamma)}{u_i(\gamma)} = \lim_{\gamma\to 0} u_i'(\gamma).$$

■

Howard and Matheson [7] also showed that under Assumptions 1-2, a finite policy improvement algorithm can be applied to find a policy $f^\infty \in \mathbf{\Pi}_{SD}$, optimal among the set of stationary deterministic policies.

5 The Average Optimality Equation

In this section we consider again a CMC with countable state space.

Theorem 1 (Average optimality equation) *: Assume that there exist $\alpha > 0$ and a bounded strictly positive function $h : \mathbf{X} \longrightarrow \mathbb{R}$ such that*

$$(sgn\gamma)\alpha h(i) = \min_{a\in\mathbf{A}} \{(sgn\gamma) e^{\gamma c(i,a)} \sum_{j\in\mathbf{X}} p_{ij}(a) h(j)\},$$

for every $i \in \mathbf{X}$. Then

$$\inf_{\pi\in\mathbf{\Pi}} \{J^\pi(i,\gamma)\} = \frac{1}{\gamma} \log \alpha,$$

for every $i \in \mathbf{X}$. Furthermore, if for each i, $f^*(i)$ attains the minimum on the right hand side of the above equation, then the policy $\pi^* = (f^*, f^*, \dots) \in \boldsymbol{Pi}_{SD}$ is optimal, that is,

$$J^{\pi^*}(i,\gamma) = J^{f^*}(i,\gamma) = \frac{1}{\gamma} \ln \alpha,$$

for every $i \in \mathbf{X}$.

Proof: Assume that the pair (α, h) satisfies the average optimality equation. Let $\pi \in \mathbf{\Pi}$ be an arbitrary policy and denote $S_0 = 0$, $S_n = \sum_{t=0}^{n-1} C_t$ for $n = 1, 2, \dots$. First, we will prove the following two statements:

i) $(\text{sgn}\gamma)\alpha^n h(i) \leqslant (\text{sgn}\gamma) E_i^\pi \left[e^{\gamma S_n} h(X_n)\right]$.

ii) $$\limsup_{n\to\infty} \frac{1}{n} \frac{1}{\gamma} \log \{E_i^\pi \left[e^{\gamma C_n} h(X_n)\right]\} = \limsup_{n\to\infty} \frac{1}{n} \frac{1}{\gamma} \log \{E_i^\pi \left[e^{\gamma S_n}\right]\}.$$

Let's prove (i) by induction on n. The validity of (i) for $n = 0$ is clear. Now, from the optimality equation we obtain that

$$(\text{sgn}\gamma)\alpha h(X_n) \leqslant (\text{sgn}\gamma) e^{\gamma C_n} E_i^\pi \left[h(X_{n+1} \mid \mathcal{F}_n\right]$$

for every $n = 0, 1, \dots$, where $\mathcal{F}_n$ is the σ-algebra generated by $\{X_0, A_0, \dots, X_n, A_n\}$. Then

$$\begin{aligned} E_i^\pi \left[E_i^\pi[(\text{sgn}\gamma)\, e^{\gamma S_n} h(X_n) \mid \mathcal{F}_{n-1}]\right] &= E_i^\pi \left[e^{\gamma S_{n-1}} (\text{sgn}\gamma) e^{\gamma C_{n-1}} E_i^\pi[h(X_n) \mid \mathcal{F}_{n-1}]\right] \\ &\geqslant E_i^\pi \left[e^{\gamma C_{n-1}} (\text{sgn}\gamma)\alpha h(X_{n-1})\right] \\ &= \alpha(\text{sgn}\gamma) E_i^\pi \left[e^{\gamma S_{n-1}} h(X_{n-1})\right] \\ &= (\text{sgn}\gamma)\alpha^n h(i) \end{aligned}$$

where the last inequality follows from the induction hypothesis. Thus, (i) follows. To prove (ii), let

$$k_1 := \sup_j \{h(j)\}, \text{ and } k_2 := \inf_j \{h(j)\}.$$

Then,

$$k_2 E_i^\pi[e^{\gamma S_n}] \leqslant E_i^\pi[h(X_n)e^{\gamma S_n}] \leqslant k_1 E_i^\pi[e^{\gamma S_n}].$$

Now, if $\gamma > 0$,

$$\begin{aligned}\frac{1}{\gamma}\log k_2 + \frac{1}{\gamma}\log E_i^\pi[e^{\gamma S_n}] \\ \leqslant \frac{1}{\gamma}\log E_i^\pi[h(X_n)e^{\gamma S_n}] \\ \leqslant \frac{1}{\gamma}\log k_1 + \frac{1}{\gamma}\log E_i^\pi[e^{\gamma S_n}]. \quad (14)\end{aligned}$$

Dividing by n, and taking lim sup, we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\frac{1}{\gamma}\log E_i^\pi[e^{\gamma S_n}] = \limsup_{n\to\infty}\frac{1}{n}\frac{1}{\gamma}\log E_i^\pi[h(X_n)e^{\gamma S_n}].$$

If $\gamma < 0$, the inequalities in (14) are just reversed and we obtain the same limiting relation, thus proving (ii).

Finally, (i) implies that

$$\frac{1}{\gamma}n\log\alpha + \frac{1}{\gamma}\log h(i) \leqslant \frac{1}{\gamma}\log E_i^\pi[h(X_n)e^{\gamma S_n}].$$

Thus, dividing by n and taking lim sup in both sides of the above equation we obtain, by (ii),

$$\frac{1}{\gamma}\log\alpha \leqslant \limsup_{n\to\infty}\frac{1}{n}\frac{1}{\gamma}\log E_i^\pi[e^{\gamma S_n}] = J^\pi(i,\gamma),$$

and therefore

$$\frac{1}{\gamma}\log\alpha \leqslant \inf_{\pi\in\Pi}\{J^\pi(i,\gamma)\}. \quad (15)$$

To prove the second part of the theorem, let f^* be the decision rule that attains the minimum in the optimality equation. By replacing A_n with $f^*(X_n)$ in the proof of (i) above, we obtain the equality in (i). Thus, a similar argument as before yields

$$\frac{1}{\gamma}\log\alpha = J^{f^*}(i,\gamma), \quad \forall i,$$

which together with (15)) implies that

$$J^{f^*}(i,\gamma) = \inf_{\pi\in\Pi}\{J^\pi(i,\gamma)\}.$$

Therefore, $\pi^* = (f^*, f^*, \ldots)$ is optimal. ■

Similarly as in the risk-neutral average case, conditions to ensure the existence of solution pairs (α, h) to the optimality equation in theorem 2 need to be explored; see [1], [5].

Acknowledgement. The authors want to thank Ms.Guadalupe Avila-Godoy, Ph.D. student in the Mathematics Department at the University of Arizona. Her M.Sc. Thesis inspired the proof of theorem 2 presented here.

References

[1] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control and Optimization*, 31(2):282–344, March 1993.

[2] E. V. Denardo and U. G. Rothblum. Optimal stopping, exponential utility, and linear programming. *Mathematical Programming*, 16:228–244, 1979.

[3] W. H. Fleming and D. Hernández-Hernández. Risk sensitive control of finite state machines on an infinite horizon i. preprint, 1995.

[4] F. R. Gantmacher. *The Theory of Matrices*, volume II. Chelsea Publishing Company, New York, N.Y., 1960.

[5] D. Hernández-Hernández and S. Marcus. Risk sensitive control of Markov processes in countable state space. Technical report, Institute for Systems Research, 1996.

[6] O. Hernández-Lerma. *Adaptive Markov Control Processes*. Springer-Verlag, New York, 1989.

[7] R. A. Howard and J. E. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, March 1972.

[8] S. C. Jaquette. Markov decision processes with a new optimality criterion: discrete time. *The Annals of Statistics*, 1:496–505, September 1973.

[9] S. C. Jaquette. A utility criterion for Markov decision processes. *Management Science*, 23(1):43–49, September 1976.

[10] T. Kato. *A Short Introduction to Perturbation Theory for Linear Operators*. Springer-Verlag, New York, 1982.

[11] S. I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernández, S. Coraluppi, and P. Fard. Risk sensitive Markov decision processes. In C. Byrnes, B. Data, D. Gilliam, and C. Martin, editors, *Systems and Control in the Twenty-First Century*, Progress in Systems and Control, pages 263–279. Birkhauser, 1996.

[12] H. Miller. A convexity property oin the theory of random variables on a finite Markov chain. *Annals of Mathematical Statistics*, 32:1260–1270, 1961.

[13] J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1):122–136, January-April 1964.

[14] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, 1994.