PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 21, Fasc. 2 (2001), pp. 493–512

RISK SENSITIVE ADAPTIVE CONTROL OF DISCRETE TIME MARKOV PROCESSES*

BY

T. E. DUNCAN, B. PASIK-DUNCAN (LAWRENCE, KANSAS)
AND Ł. STETTNER** (WARSZAWA)

Abstract. Adaptive control of discrete time Markov processes with an infinite horizon risk sensitive cost functional is investigated. The continuity of the optimal risk sensitive cost with respect to a parameter of the transition probability is verified. Two almost optimal adaptive procedures that are based on the large deviations of the cost functional and discretized maximum likelihood estimates are given. To justify the performance of the adaptive procedure with observations of the cost, some large deviations estimates of the empirical distributions of finite sequences of successive states of Markov processes are obtained. A finite family of continuous control functions, where one control function is fixed after a nonrandom time from each of the adaptive procedures, provides an almost optimal adaptive control.

AMS Subject Classification: 93C40, 93E12, 93E20, 60J05.

Key words and phrases: Adaptive control, risk sensitive control, discrete time controlled Markov processes, ergodic control.

1. INTRODUCTION

Let $(X_n, n \in N)$ be a controlled, E-valued Markov process on the probability space $(\Omega, \mathscr{F}, P)$, where E is a locally compact, separable metric space and $\mathscr{E}$ is the Borel σ-algebra on E. The process $(X_n, n \in N)$ has a controlled transition probability $P^{\alpha^0 v_n}(X_n, dy)$, where α^0 is an unknown parameter that takes values in a compact metric space A, and v_n is a control at time n that is adapted to $\sigma(X_1, \ldots, X_n)$ and takes values in a compact metric space U.

* Research supported in part by NSF Grant DMS 9623439, KBN Grant 2 P03A 0509 and KBN Grant 2 P03A 015 15.

** Institute of Mathematics, Polish Academy of Sciences, Warszawa.

Let $c\colon E\times U\to \boldsymbol{R}_+$ be a continuous bounded function. Let λ_α be the infimum of the risk sensitive infinite horizon cost functional

$$(1)\qquad J_x^\alpha((v_n,\ n\in N)) = \limsup_{n\to\infty}\frac{1}{n}\log E_x^\alpha[\exp(\sum_{i=0}^{n-1}\gamma c(X_i,\ v_i))]$$

over the family of $\sigma\{X_1, \ldots, X_n\}$-adapted U-valued controls, given value of the parameter $\alpha\in A$ in the transition probability of the process $(X_n,\ n\in N)$ and a positive risk factor γ. The term E_x^α in (1) is used to denote conditional expectation given the initial state $X_0 = x$ and the value of the parameter α. Under the assumptions that are imposed, it follows that the optimal cost λ_α does not depend on the initial state x. Since the transition probability measures of the Markov process depend on the unknown parameter $\alpha\in A$, an adaptive procedure is required to construct a nearly optimal control.

In the paper the following problem is solved:

For a given $\varepsilon > 0$ find a nonrandom time T and an adaptive control $(v_n, n\in N)$ such that

$$(2)\qquad \limsup_{n\to\infty} E_x^{\alpha^0}\left[\frac{1}{n}\log E^{\alpha^0}[\exp(\sum_{i=0}^{n-1}\gamma c(X_i,\ v_i))\mid X_1, \ldots, X_T]\right]\leqslant \lambda_{\alpha^0}+\varepsilon,$$

where λ_{α^0} is the infimum of (1) and $\alpha^0\in A$ is unknown. Thus an adaptive strategy is sought that starting from a nonrandom time T almost minimizes the expected value of the risk sensitive cost conditional on $\sigma(X_1, \ldots, X_T)$.

Two approaches to adaptive control are used: the first is based on observation of the cost and the second is based on discretized maximum likelihood estimates that are motivated by [6] and [7].

In Section 2, continuity of the solution of the Hamilton–Jacobi–Bellman equation [2] with respect to the (unknown) parameter for the risk sensitive control problem is verified. These continuity results are similar to those of [1] and [7] for average cost per unit time control problems. In Section 3, some large deviations estimates for finite sequences of successive states of the Markov process are verified. These estimates are used to construct an adaptive, almost optimal control based on observation of the cost. In Section 4, an adaptive control is constructed from a family of discretized maximum likelihood estimates [6]. Similar adaptive controls are constructed in [7] for average cost per unit time control problems. The main contribution of this paper is the construction of an almost optimal adaptive control for risk sensitive cost functionals. It seems that this is the first work on adaptive control for risk sensitive cost functionals.

2. RISK SENSITIVE CONTROL PROBLEMS DEPENDING ON A PARAMETER

The transition measures for the controlled Markov process are assumed to have densities with respect to a fixed probability measure, that is, for $x \in E$, $\alpha \in A$, $v \in U$, $B \in \mathscr{E}$,

$$P^{\alpha v}(x, B) = \int_B p(x, y, \alpha, v)\,\eta(dy), \tag{3}$$

where η is a fixed probability measure on E. The following assumptions are made concerning the density p:

(A1) The density p: $E \times E \times A \times U \to \boldsymbol{R}_+$ is continuous and bounded and, for each x, $y \in E$, $\alpha \in A$, and $v \in U$, $p(x, y, \alpha, v) > 0$.

(A2) $\sup_{\alpha\in A}\sup_{x,x'\in E}\sup_{y\in E}\sup_{v,v'\in U}\dfrac{p(x, y, \alpha, v)}{p(x', y, \alpha, v')} := K < \infty.$

(A3) $\exists\delta < 1$ such that $\forall x, x' \in E$, $\forall \alpha \in A$, $\forall v, v' \in U$, $\forall B \in \mathscr{E}$,

$$P^{\alpha v}(x, B) - P^{\alpha v'}(x', B) < \delta$$

and $\delta \exp(\gamma \|c\|_{\text{sp}}) < 1$, where $\|c\|_{\text{sp}} = \sup_{x,v} c(x, v) - \inf_{x',v'} c(x', v')$.

Remark. The assumptions (A1) and (A2) are satisfied when E is compact and p: $E \times E \times A \times U \to \boldsymbol{R}_+$ is continuous. If $E = \boldsymbol{R}^d$ and

$$X_{n+1} = f(X_n, \alpha^0, v_n) + g(X_n) W_n,$$

where f: $\boldsymbol{R}^d \times A \times U \to \boldsymbol{R}^d$ are g: $\boldsymbol{R}^d \to \mathscr{L}(\boldsymbol{R}^d)$ are continuous and bounded, g^{-1}: $\boldsymbol{R}^d \to \mathscr{L}(\boldsymbol{R}^d)$ and $(W_n, n \in \boldsymbol{N})$ is a sequence of $\boldsymbol{R}^d$-valued independent identically distributed standard Gaussian random variables, then (A1) and the first part of (A3) are satisfied and, consequently, for γ sufficiently small also $\delta \exp(\gamma \|c\|_{\text{sp}}) < 1$.

Let g: $E \to \boldsymbol{R}$ be a Borel measurable function and define T^α as

$$(T^\alpha g)(x) := \inf_{v\in U}\left[\gamma c(x, v) + \log \int_E e^{g(y)} P^{\alpha v}(x, dy)\right]. \tag{4}$$

Let $C_L(E)$ be the subspace of the space $C(E)$ of bounded, continuous functions on E whose span norm $\|\cdot\|_{\text{sp}}$ is bounded by L, that is, $f \in C_L(E)$ if

$$\|f\|_{\text{sp}} = \sup_{x\in E} f(x) - \inf_{y\in E} f(y) \leqslant L.$$

The following result gives an important property of T^α acting on $C_L(E)$.

THEOREM 1. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied, for any* $L > 0$ *the operator* T^α: $C_L(E) \to C(E)$ *given by* (4) *is a span norm contraction. There is a unique pair* $(\lambda_\alpha, w_\alpha) \in \boldsymbol{R}_+ \times C_L(E)$ *such that, for each* $x \in E$ *and a fixed* $\bar{x} \in E$,

$$T^\alpha w_\alpha(x) = \lambda_\alpha + w_\alpha(x), \qquad w_\alpha(\bar{x}) = 0, \tag{5}$$

where, under (A1) *and* (A2), $L = \log K + \|c\|_{sp}$ *with* K *given in* (A2), *while under* (A1) *and* (A3),

$$L = \gamma \|c\|_{sp} + \ln \left[\sum_{i=0}^{\infty} (\delta \exp(\gamma \|c\|_{sp}))^i \right].$$

Proof. Note first that by (A2) there is a $\delta < 1$ such that

$$P^{\alpha v}(x, B) - P^{\alpha v'}(x', B) \leqslant \delta \tag{6}$$

for all $\alpha \in A$, $x, x' \in E$, $v, v' \in U$ and $B \in \mathscr{E}$. Following the proof of Proposition 2 in [2] it follows that for $g_1, g_2 \in C(E)$

$$\|T^\alpha g_1 - T^\alpha g_2\|_{sp} \leqslant \|g_1 - g_2\|_{sp} \sup_{B \in \mathscr{E}} \sup_{x, x' \in E} \sup_{v, v' \in U} \left(\mu_{xv\alpha g_1}(B) - \mu_{x'v'\alpha g_2}(B)\right), \tag{7}$$

where

$$\mu_{xv\alpha g}(B) = \frac{\int_B e^{g(y)} P^{\alpha v}(x, dy)}{\int_E e^{g(y)} P^{\alpha v}(x, dy)}.$$

To complete the proof of the (span norm) contraction property it suffices to show that

$$\sup_{\alpha \in A} \sup_{g_1, g_2, \|g_1\|_{sp} \leqslant M, \|g_2\|_{sp} \leqslant M} \sup_{x, x' \in E} \sup_{v, v' \in U} \sup_{B \in \mathscr{E}} \left(\mu_{xv\alpha g_1}(B) - \mu_{x'v\alpha g_2}(B)\right) := \beta(M) < 1. \tag{8}$$

Following the proof of Proposition 2 in [2] it can be shown that if (8) is not satisfied, then (6) is not satisfied. To complete the proof use Theorem 1 of [2] in the case of (A1) and (A2) and Proposition 2 and Theorem 1 of [3] in the case of (A1) and (A3). ∎

Corollary 1. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied, then*

$$\lim_{\alpha_n \to \alpha} \|w_{\alpha_n} - w_\alpha\| = 0.$$

Proof. By Theorem 1 of [2], under (A1) and (A2) the operator T^α transforms $C_L(E)$ with L given in Theorem 1 into itself. Under (A1) and (A3), by Proposition 2 of [3] the iterations of $T^\alpha 0$ belong to $C_L(E)$ with L as in Theorem 1. Using a contraction principle approximation with contraction constant $\beta(L)$ we obtain

$$\begin{aligned} \|w_{\alpha_n} - w_\alpha\|_{sp} &\leqslant \|w_{\alpha_n} - (T^{\alpha_n})^m(0)\|_{sp} + \|(T^{\alpha_n})^m(0) - (T^\alpha)^m(0)\|_{sp} \\ &\leqslant 2(\beta(L))^m L + \|(T^{\alpha_n})^m(0) - (T^\alpha)^m(0)\|_{sp}. \end{aligned}$$

For each $m \in N$ from $\|(T^{\alpha_n})^m(0) - (T^\alpha)^m(0)\|_{sp} \to 0$ as $n \to \infty$ it follows that

$$\lim_{n \to \infty} \|w_{\alpha_n} - w_\alpha\|_{sp} = 0.$$

Therefore, there is a sequence of real numbers $(d_n, n \in N)$ such that

$$\lim_{n \to \infty} \|w_{\alpha_n} - w_\alpha - d_n\| = 0.$$

Since $w_{\alpha_n}(\bar{x}) = w_\alpha(\bar{x}) = 0$, it follows that $d_n \to 0$ as $n \to \infty$. ■

COROLLARY 2. *If* (A2) *and* (A2) *or* (A1) *and* (A3) *are satisfied, then*

$$\lim_{\alpha_n \to \alpha} \lambda_{\alpha_n} = \lambda_\alpha.$$

Proof. Note that

$$\lambda_{\alpha_n} = \inf_{v \in U} \left[\gamma c(x, v) + \log \int_E \exp\left(w_{\alpha_n}(y)\right) P^{\alpha_n v}(x, dy)\right] - w_{\alpha_n}(x).$$

It follows from Corollary 1 and (A1) that $\lambda_{\alpha_n} \to \lambda_\alpha$ as $n \to \infty$. ■

Let u: $E \to U$ be Borel measurable and for each $\alpha \in A$ define λ_α^u as

$$\lambda_\alpha^u = \limsup_{n \to \infty} \frac{1}{n} \log E_x^{\alpha u}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, v_i)\right)\right], \tag{9}$$

where $v_i = u(X_i)$.

The next lemma follows from Proposition 2 of [2] (under the assumptions (A1) and (A3) using the arguments of the proof of Theorem 1 we obtain easily an analog of Proposition 2 of [2]) and the proof of Corollary 2 of [2].

LEMMA 1. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied, then for* $n \in N$

$$-L \leqslant \log E_x^{\alpha u}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\right)\right] - n\lambda_\alpha^u \leqslant L, \tag{10}$$

where L *is given in Theorem* 1 *and* u *is Borel measurable.*

The following result proves the existence of a finite family of almost optimal controls.

PROPOSITION 1. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied, then for* $\varepsilon > 0$ *there is a finite family* $\mathcal{U}(\varepsilon) = \{u_1, \ldots, u_k\}$ *of* ε*-optimal controls, that is,* u_m: $E \to U$ *is Borel measurable for* $m = 1, 2, \ldots, k$ *and for each* $\alpha \in A$ *there is an* $m \in \{1, \ldots, k\}$ *such that*

$$J_x^\alpha\left((u_m(X_n), n \in N)\right) \leqslant \lambda_\alpha + \varepsilon.$$

Proof. By Propositions 1 of [2] and [3] and Theorems 1 of [2] and [3] for each $\alpha \in A$ there is an optimal control u_α. Let $n_0 \in N$ be such that $L/n_0 \leqslant \varepsilon/4$. By Lemma 1 for u: $E \to U$ Borel measurable and $\alpha, \alpha' \in A$ it follows that for $n \geqslant n_0$

$$|\lambda_{\alpha'}^u - \lambda_\alpha^u| \leqslant \frac{2L}{n} + \frac{1}{n}\left|\log \frac{E_x^{\alpha' u}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\right)\right]}{E_x^{\alpha u}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\right)\right]}\right|$$

$$\leqslant \frac{\varepsilon}{2} + \frac{1}{n} F_n(\alpha', \alpha, x, u).$$

If $F_n(\alpha', \alpha, x, u) \to 0$ as $\alpha' \to \alpha$, then

$$\lambda^u_\cdot : A \to \boldsymbol{R}_+$$

is continuous for all Borel measurable $u: E \to U$. Thus the family $\{\alpha' \in A: |\lambda^{u_\alpha}_{\alpha'} - \lambda^{u_\alpha}_{\alpha}| < \varepsilon\}$, for $\alpha \in A$ and u_α Borel measurable, is an open covering of the compact set A. Consequently,

$$A \subset \bigcup_{\alpha \in \Lambda} \{\alpha' \in A: |\lambda^{u_\alpha}_{\alpha'} - \lambda^{u_\alpha}_{\alpha}| < \varepsilon\},$$

where Λ is a finite set, so for $\mathscr{U}(\varepsilon) = \{u_\alpha, \alpha \in \Lambda\}$ the assertion of Proposition 1 is satisfied.

It only remains to show that $F_n(\alpha', \alpha, x, u) \to 0$ as $\alpha' \to \alpha$. Initially, it is shown that

$$(11) \qquad \lim_{m \to \infty} E^{\alpha_m}_x \Big[\exp\Big(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\Big) \varphi_m(X_n)\Big] = E^\alpha_x \Big[\exp\Big(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\Big) \varphi(X_n)\Big],$$

where $\alpha_m \to \alpha$ as $m \to \infty$ and $(\varphi_m, m \in N)$ is a family of uniformly bounded, Borel measurable functions that converge uniformly on compact subsets of E to a function φ. Given $\varepsilon > 0$, for $n = 1$ choose a compact set K such that $\eta(K) \geqslant 1 - \varepsilon$ and

$$(12) \qquad \begin{aligned} &|e^{\gamma c(x,u(x))} E^{\alpha_m}_x [\varphi_m(X_1)] - e^{\gamma c(x,u(x))} E^{\alpha_m}_x [\varphi(X_1)]| \\ &\leqslant e^{\gamma \|c\|} \int_E |\varphi(y)| \, |p(x, y, \alpha_m, u(x)) - p(x, y, \alpha, u(x))| \, \eta(dy) \\ &\leqslant 2e^{\gamma \|c\|} \|\varphi\| \, \|p\| \, \varepsilon + e^{\gamma \|c\|} \|\varphi\| \sup_{y \in K} |p(x, y, \alpha_m, v(x)) - p(x, y, \alpha, u(x))|. \end{aligned}$$

By (A1), the right-hand side of (12) can be made arbitrarily small. Furthermore, it is clear that by (12)

$$\lim_{m \to \infty} e^{\gamma c(y,u(y))} E^{\alpha_m}_y [\varphi_m(X_1)] = e^{\gamma c(y,u(y))} E^\alpha_y [\varphi(X_1)]$$

uniformly in y from compact subsets of E. Assuming by induction that (11) is true for n, it is verified for $n+1$, that is,

$$E^{\alpha_m}_x \Big[\exp\Big(\sum_{i=0}^{n} \gamma c(X_i, u(X_i))\Big) \varphi_m(X_{n+1})\Big] = E^{\alpha_m}_x \Big[\exp\Big(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\Big) \exp\big(\gamma c(X_n, u(X_n))\big) E^{\alpha_m}_{X_n} [\varphi_m(X_1)]\Big]$$

$$\to E_x^\alpha\big[\exp\big(\sum_{i=0}^{n-1}\gamma c(X_i, u(X_i))\big)\exp\big(\gamma c(X_n, u(X_n))\big)E_{X_n}^\alpha[\varphi(X_1)]\big]$$

$$= E_x^\alpha\big[\exp\big(\sum_{i=0}^{n}\gamma c(X_i, u(X_i))\big)\varphi(X_{n+1})\big] \quad \text{as } m\to\infty.$$

This verifies the induction hypothesis. Letting $\varphi_m = \varphi \equiv 1$ we have

$$\lim_{m\to\infty} E_x^{\alpha_m}\big[\exp\big(\sum_{i=0}^{n-1}\gamma c(X_i, u(X_i))\big)\big] = E_x^\alpha\big[\exp\big(\sum_{i=0}^{n-1}\gamma c(X_i, u(X_i))\big)\big],$$

whence we obtain

$$\lim_{m\to\infty} F_n(\alpha_m, \alpha, x, u) = 0. \;\blacksquare$$

COROLLARY 3. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied for each Borel measurable function* $u\colon E\to U$, *then the mapping* $\lambda^u_\cdot\colon A\to R_+$ *is continuous, where* λ^u_α *is given by* (9).

This corollary follows immediately from the arguments used in the proof of Proposition 1.

The main result of this section is the following result on the existence of a finite family of continuous almost optimal controls.

THEOREM 2. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied, then for each* $\varepsilon>0$ *there is a finite family* $\mathcal{U}_c(\varepsilon) = \{u_1, \ldots, u_k\}$ *of* ε*-optimal continuous control functions, that is, for each* $\alpha\in A$ *there is an* $m\in\{1,\ldots,k\}$ *such that*

$$J_x^\alpha\big((u_m(X_n), n\in N)\big) \leqslant \lambda_\alpha+\varepsilon, \tag{13}$$

and $u_j\colon E\to U$ *for* $j\in\{1,\ldots,k\}$ *is continuous.*

Proof. Since E is locally compact, there is an increasing sequence $(K_n, n\in N)$ of compact sets such that $\bigcup_{n=1}^\infty K_n = E$. It can be assumed that $\eta(\partial K_n) = 0$ for all $n\in N$. Consider a sequence of partitions $((E_1^n, \ldots, E_{d_n}^n), n\in N)$ of E with representative elements $((e_1^n, \ldots, e_{d_n}^n), n\in N)$ such that $e_i^j\in E_i^j$, $E = \bigcup_{i=1}^{d_n} E_i^n$, $E_i^n\cap E_j^n = \emptyset$ if $i\neq j$, $\eta(\partial E_i^n) = 0$, the diameter of E_i^n is not greater than $1/n$ for $j\in\{1,\ldots,d_n-1\}$, $E_{d_n}^n = E\backslash K_n$, $(E_1^{n+1}, \ldots, E_{d_{n+1}}^{n+1})$ is a subpartition of $(E_1^n, \ldots, E_{d_n}^n)$, and $\{e_1^n, \ldots, e_{d_n}^n\}\subset\{e_1^{n+1}, \ldots, e_{d_{n+1}}^{n+1}\}$ for each $n\in N$.

Let $p_n^{\alpha v}(e_i^n, e_j^n) = P^{\alpha v}(e_i^n, E_j^n)$. Consider the controlled Markov process with the state space $\{e_1^n, \ldots, e_{d_n}^n\}$ and the transition probabilities $p_n^{\alpha v}(e_i^n, e_j^n)$ for $i, j\in\{1,\ldots,d_n\}$. The mapping

$$p_n^{\cdot\cdot}(e_i, e_j)\colon A\times U\to R_+$$

is continuous and under (A2) we obtain

$$\sup_n\sup_{\alpha\in A}\sup_{i,j\in\{1,\ldots,d_n\}}\sup_{v,v'\in U}\frac{p_n^{\alpha v}(e_i^n, e_j^n)}{p_n^{\alpha v'}(e_i^n, e_j^n)} \leqslant K, \tag{14}$$

while under (A3) we have

$$p_n^{\alpha v}(e_i^n, e_j^n) - p_n^{\alpha v'}(e_i^n, e_j^n) < \delta$$

with the same constants K and δ as in (A2) and (A3).

Proposition 1 applied to the Markov process with state space $(e_1^n, \ldots, e_{d_n}^n)$ implies that there is a finite family $(\bar{u}_1^n, \ldots, \bar{u}_{k_n}^n)$ of $(\varepsilon/3)$-optimal control functions. Now let $u_j^n(x) = \bar{u}_j^n(e_i^n)$ and $P_n^{\alpha v}(x, \cdot) = P^{\alpha v}(e_i^n, \cdot)$ for $x \in E_i^n$, $i \in \{1, \ldots, d_n\}$ and $n \in N$. It is clear that for each fixed $m \in N$ the family of piecewise constant functions $(u_1^m, \ldots, u_{k_m}^m)$ is $(\varepsilon/3)$-optimal for the Markov process $(X_n^m, n \in N)$ with the transition measure $P_m^{\alpha v}(x, \cdot)$. For a Borel measurable function $u\colon E \to U$ let $\lambda_{\alpha,m}^u$ be given by

$$\lambda_{\alpha,m}^u = \limsup_{n \to \infty} \frac{1}{n} \log E_x^\alpha \Big[\exp\Big(\sum_{i=0}^{n-1} \gamma c(X_i^m, u(X_i^m))\Big)\Big]. \tag{15}$$

From Lemma 1 it follows that for $n \in N$

$$-L \leqslant \log E_x^\alpha \Big[\exp\Big(\sum_{i=0}^{n-1} \gamma c(X_i^m, u(X_i^m))\Big)\Big] - n\lambda_{\alpha,m}^u \leqslant L. \tag{16}$$

Therefore,

$$\begin{aligned} \sup_{\alpha \in A} \sup_u |\lambda_{\alpha,m}^u - \lambda_\alpha^u| \\ \leqslant \frac{2L}{n} + \sup_{\alpha \in A} \sup_u \frac{1}{n} \left| \log \frac{E_x^\alpha \big[\exp\big(\sum_{i=0}^{n-1} \gamma c(X_i^m, u(X_i^m))\big)\big]}{E_x^\alpha \big[\exp\big(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\big)\big]} \right|, \end{aligned} \tag{17}$$

where the supremum over u is over the family of Borel measurable functions $u\colon E \to U$.

Since

$$\lim_{m \to \infty} \sup_{\alpha \in A} \sup_{v \in U} \|P_m^{\alpha v}(x, \cdot) - P^{\alpha v}(x, \cdot)\|_{\mathrm{var}} = 0, \tag{18}$$

where the convergence is uniform in x from compact sets and $\|\cdot\|_{\mathrm{var}}$ denotes the variation norm. Choosing n sufficiently large in (17) and using (18), for $m \geqslant m_0$ we obtain

$$\sup_{\alpha \in A} \sup_u |\lambda_{\alpha,m}^u - \lambda_\alpha^u| < \varepsilon/3.$$

Thus the family of piecewise constant functions $(u_1^m, \ldots, u_{k_m}^m)$ is $(2\varepsilon/3)$-optimal for the Markov process $(X_n, n \in N)$ with the transition measure $P^{\alpha v}(x, \cdot)$.

It only remains to show that each $u \in \{u_1^m, \ldots, u_{k_m}^m\}$ can be approximated by a sequence $(u_l, l \in N)$ of continuous control functions such that

$$\lim_{l \to \infty} \sup_{\alpha \in A} |\lambda_\alpha^{u_l} - \lambda_\alpha^u| = 0. \tag{19}$$

Given a $u \in \{u_1^m, \ldots, u_{k_m}^m\}$, choose a sequence $(u_l, l \in N)$ of continuous U-valued functions on E such that

$$\lim_{l \to \infty} \eta(u_l \neq u) = 0.$$

Thus,

$$\lim_{l \to \infty} \eta\big(z\colon\; p(z, y, \alpha, u_l(z)) \neq p(z, y, \alpha, u(z)) \text{ for some } \alpha \in A,\; y \in E\big) = 0$$

and

$$\lim_{l \to \infty} \eta\big(z\colon\; \sup_{\alpha \in A} \|P^{\alpha u_l(z)}(z, \cdot) - P^{\alpha u(z)}(z, \cdot)\|_{\text{var}} > 0\big) = 0.$$

Hence for each u

$$\lim_{\substack{n \to \infty \\ l \to \infty}} \sup_{\alpha \in A} \left| \log \frac{E_x^{\alpha u_l}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u_l(X_i))\right)\right]}{E_x^{\alpha u}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\right)\right]} \right| = 0 \;\; \eta \text{ a.e.} \tag{20}$$

By Lemma 1 it follows that for $n \in N$

$$\sup_{\alpha \in A} |\lambda_\alpha^{u_l} - \lambda_\alpha^u| \leqslant \frac{2L}{n} + \sup_{\alpha \in A} \left| \log \frac{E_x^{\alpha u_l}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u_l(X_i))\right)\right]}{E_x^{\alpha u}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, u(X_i))\right)\right]} \right|.$$

So choosing n and l sufficiently large it follows by (20) that

$$\sup_{\alpha \in A} |\lambda_\alpha^{u_l} - \lambda_\alpha^u| \leqslant \varepsilon/3.$$

Thus there is a finite family of continuous ε-optimal control functions. ∎

To end this section a continuity result for the invariant measures of the Markov process with respect to parameters is given.

LEMMA 2. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied and* $u\colon E \to U$ *is continuous, then the mapping*

$$\pi^u\colon A \to \mathscr{P}(E)$$

is continuous in the variation norm, where π_α^u *is the unique invariant measure for the transition operator* $P^{\alpha u}(x, \cdot)$.

Proof. By (6) and (A3) using Lemma 3.3 of [8] we see that for each Borel measurable $u\colon E \to U$ there is a unique invariant measure π_α^u and for $x \in E$

$$\|(P^{\alpha v})^n(x, \cdot) - \pi_\alpha^u\|_{\text{var}} \leqslant 2\delta^n \quad \text{ for } \alpha \in A.$$

Since $\delta < 1$ does not depend on u and α, the family of invariant measures $(\pi_\alpha^u, \alpha \in A)$ is tight.

Let now $u\colon E \to U$ be continuous. If $\alpha_n \to \alpha$ as $n \to \infty$, then there is a subsequence, denoted as the whole sequence for notational simplicity, such that

$\pi^u_{\alpha_n} \Rightarrow \mu$, where μ is a probability measure on E. If $f\colon E \to \boldsymbol{R}$ is continuous and bounded, then, taking into account that $\pi^u_{\alpha_n}(P^{\alpha_n u} f) = \pi^u_{\alpha_n}(f)$, we have

$$|\mu(P^{\alpha u} f) - \mu(f)| \leqslant |\mu(P^{\alpha u} f) - \pi^u_{\alpha_n}(P^{\alpha u} f)| \\ + |\pi^u_{\alpha_n}(P^{\alpha u} f) - \pi^u_{\alpha_n}(P^{\alpha_n u} f)| + |\pi^u_{\alpha_n}(f) - \mu(f)|.$$

Since

$$|\pi^u_{\alpha_n}(P^{\alpha u} f) - \pi^u_{\alpha_n}(P^{\alpha_n u} f)| \leqslant \sup_{x \in K} |P^{\alpha u} f(x) - P^{\alpha_n u} f(x)| + 2\,\|f\|\, \pi^u_{\alpha_n}(K^c)$$

and, for $\varepsilon > 0$, $\pi^u_{\alpha_n}(K^c) \leqslant \varepsilon/2\,\|f\|$ for all $n \in N$ for a suitable compact set K, it follows that $\mu(P^{\alpha u} f) = \mu(f)$ for all bounded continuous functions f. Thus μ is an invariant measure and, by uniqueness, $\mu = \pi^u_\alpha$ and the mapping

$$\pi^u_\cdot\colon A \to \mathscr{P}(E)$$

is continuous in the weak* topology.

For $B \in \mathscr{E}$ and a sequence $(\alpha_n, n \in N)$ such that $\alpha_n \to \alpha$ as $n \to \infty$ it follows that

$$|\pi^u_\alpha(B) - \pi^u_{\alpha_n}(B)| \leqslant \pi^u_{\alpha_n}(K^c) + \sup_{x \in K} |P^{\alpha u}(x, B) - P^{\alpha_n u}(x, B)| \\ + \Big|\int_E \int_B p\big(x, y, \alpha, u(x)\big)\, \eta(dy) \big(\pi^u_\alpha(dx) - \pi^u_{\alpha_n}(dx)\big)\Big|.$$

Since

$$\Big|\int_E \int_B p\big(x, y, \alpha, u(x)\big)\, \eta(dy) \big(\pi^u_\alpha(dx) - \pi^u_{\alpha_n}(dx)\big)\Big| \\ \leqslant \int_E \Big|\int_E p\big(x, y, \alpha, u(x)\big) \big(\pi^u_\alpha(dx) - \pi^u_{\alpha_n}(dx)\big)\Big|\, \eta(dy)$$

and the right-hand side tends to zero as $n \to \infty$ by the weak* convergence of $(\pi^u_{\alpha_n}, n \in N)$ to π^u_α, continuity of u, it follows that $(\pi^u_{\alpha_n}, n \in N)$ converges to π^u_α in the variation norm. ∎

3. ADAPTIVE CONTROL WITH OBSERVATION OF THE COST

Initially some results on the large deviations of empirical distributions of finite sequences of the states of the Markov process are given. These results are used for the adaptive controls.

In this section $(X_n, n \in N)$ is a Markov process with the transition probability measure $P^\alpha(x, \cdot)$ depending on a parameter $\alpha \in A$. The following assumption is used:

(B1) For each $f \in C(E)$, the mapping

$$A \times E \ni (\alpha, x) \mapsto P^{\alpha} f(x) = \int_E f(y) P^{\alpha}(x, dy)$$

is continuous.

Fix $N \in N$ and for $B \in \mathscr{B}(E^N)$ and $n \in N$ define the empirical distribution S_n^N as

$$S_n^N(B) = \frac{1}{n} \sum_{j=0}^{n-1} 1_B(X_{jN+1}, X_{jN+2}, \ldots, X_{(j+1)N}). \tag{21}$$

Some asymptotic properties of the measure

$$Q_{nx}^{N,\alpha,u}(\Gamma) = P_x^{\alpha u}\{S_n^N \in \Gamma\} \tag{22}$$

for $\Gamma \in \mathscr{B}(\mathscr{P}(E^N))$ are given.

Let $\bar{x} = (x_1, \ldots, x_N)$ and $\bar{y} = (y_1, \ldots, y_N)$ be two vectors in E^N. Let Φ be given as

$$\Phi = \{f \in C(E^N)\colon \text{ there is } a > 0 \text{ such that } f(\bar{x}) \geqslant a \text{ for all } \bar{x} \in E^N\}$$

and for $l \in \mathscr{P}(E^N)$ let I^{α} be given as

$$I^{\alpha}(l) = \sup_{f \in \Phi} \int_{E^N} \log\left(\frac{f(y_1, \ldots, y_N)}{E_{y_N}^{\alpha}[f(X_1, \ldots, X_N)]}\right) l(dy_1, \ldots, dy_N). \tag{23}$$

I^{α} is related to $\log Q$ in the following result.

THEOREM 3. *If* (B1) *is satisfied, then for compact sets* $C \subset \mathscr{P}(E^N)$ *and* $A_1 \subset A$ *and a positive integer* N

$$\limsup_{n \to \infty} \frac{1}{n} \sup_{\alpha \in A_1} \sup_{x \in E} \log Q_{nx}^{N,\alpha,u}(C) \leqslant -\inf_{\alpha \in A_1} \inf_{l \in C} I^{\alpha}(C). \tag{24}$$

Proof. For $d > 0$ let

$$\Phi_d = \{f \in \Phi\colon \sup_{\bar{x} \in E^N} f(\bar{x}) \leqslant d \inf_{\bar{y} \in E^N} f(\bar{y})\}.$$

Following the proof of Theorem 1 in [5] it can be shown that for $f \in \Phi_d$

$$E_x^{\alpha}\left[\exp\left(\sum_{j=0}^{n-1} \log \frac{f(X_{jN+1}, \ldots, X_{(j+1)N})}{E_{X_{(j+1)N}}^{\alpha}[f(X_1, \ldots, X_N)]}\right)\right] \leqslant d,$$

and therefore for $\Gamma \in \mathscr{B}(\mathscr{P}(E^N))$

$$Q_{nx}^{N,\alpha,u}(\Gamma) \leqslant d \exp\left[-n \inf_{l \in \Gamma} \int_{E^N} \log \frac{f(y_1, \ldots, y_N)}{E_{y_N}^{\alpha}[f(X_1, \ldots, X_N)]} l(dy_1, \ldots, dy_N)\right].$$

The remaining part of the proof follows as in Theorems 1 and 4 in [7]. ■

To obtain a stronger version of Theorem 3 for closed sets C an analogue of the assumption (B4) of [7] is used.

(B2) There is a continuous function $\psi: A \times E^N \to R$ such that $\psi(\alpha, \bar{x}) \geqslant 1$ for $\alpha \in A$, $\bar{x} \in E^N$, and the mapping

$$\sup_{\alpha \in A} E^{\alpha}_{\cdot} [\psi(\alpha, X_1, \ldots, X_N)]: E \to R$$

is bounded on compact subsets of E and for any $m > 0$ the set K_m given by

$$K_m = \left\{ (y_1, \ldots, y_N) \in E^N: \inf_{\alpha \in A} \frac{\psi(\alpha, y_1, \ldots, y_N)}{E^{\alpha}_{y_N} [\psi(\alpha, X_1, \ldots, X_N)]} \leqslant m \right\}$$

is compact in E^N.

An analysis of the proofs of Lemmas 4 and 5 and Theorem 5 of [7] shows that the following analogue of Theorem 5 of [7] is satisfied.

THEOREM 4. *If* (B1) *and* (B2) *are satisfied for closed sets* $C \subset \mathscr{P}(E^N)$ *and* $A_1 \subset A$, *then for any compact subset* $W \subset E$

$$\limsup_{n \to \infty} n^{-1} \sup_{\alpha \in A_1} \sup_{x \in W} \log Q^{N,\alpha,u}_{nx}(C) \leqslant - \inf_{\alpha \in A_1} \inf_{l \in C} I^{\alpha}(C). \tag{25}$$

Following the proof of Lemma 4.2 in [5] an analogue of Lemma 6 of [7] is obtained.

LEMMA 3. *If* (B1) *and* (B2) *are satisfied, then for* $m > 0$ *the set*

$$C_m = \{ l \in \mathscr{P}(E^N): \inf_{\alpha \in A} I^{\alpha}(l) \leqslant m \}$$

is compact.

The following property of I^{α} is useful.

LEMMA 4. *For* $\alpha \in A$ *and* $l \in \mathscr{P}(E^N)$, $I^{\alpha}(l) = 0$ *if and only if for each* $B \in \mathscr{B}(E^N)$

$$l(B) = \int_E E^{\alpha}_x [1_B(X_1, \ldots, X_N)] \, \pi_N(dx),$$

where π_N *is an invariant measure for the Markov process* $(X_{iN}, i \in N)$ *with the transition operator* $(P^{\alpha})^N(x, \cdot)$.

Proof. Following the proof of Lemma 2.5 in [4] it suffices to show that $I^{\alpha}(l) = 0$ is equivalent to

$$\int_{E^N} f(y_1, \ldots, y_N) \, l(dy_1, \ldots, dy_N) = \int_E E^{\alpha}_x [f(X_1, \ldots, X_N)] \, l(E^{N-1}, dx)$$

for $f \in \Phi$. ■

An adaptive control procedure is described for a given $\varepsilon > 0$ and an initial state $x \in E$ such that in a nonrandom time T the inequality (2) is satisfied.

The procedure consists of the following five steps:

1. Determine the finite family $\mathscr{U}_c(\varepsilon/4) = \{u_1, \ldots, u_k\}$ of continuous $(\varepsilon/4)$-optimal control functions as given in Theorem 2.

2. Determine a positive integer N such that for $x' \in E$, $\alpha \in A$, and $u \in \mathscr{U}_c(\varepsilon/4)$

$$-\frac{\varepsilon}{8} \leqslant \frac{1}{N} \log E_{x'}^{\alpha u}\left[\exp\left(\sum_{i=1}^{N} \gamma c(X_i, u(X_i))\right)\right] - \lambda_\alpha^u \leqslant \frac{\varepsilon}{8}. \tag{26}$$

3. Determine a positive integer n such that for $x' \in E$, $\alpha \in A$, and $u \in \mathscr{U}_c(\varepsilon/4)$

$$P_{x'}^{\alpha u}\left[\left|\frac{1}{n}\sum_{j=0}^{n-1} \exp\left(\sum_{i=jN+1}^{(j+1)N} \gamma c(X_i, u(X_i))\right) - \int_E E_y^{\alpha u}\left[\exp\left(\sum_{i=1}^{N} \gamma c(X_i, u(X_i))\right)\right]\pi_\alpha^u(dy)\right| \geqslant \frac{\varepsilon}{8}\right] \leqslant \frac{\varepsilon}{4\|c\|\,k}, \tag{27}$$

where π_α^u is an invariant measure of the Markov process with transition operator $P^{\alpha u}(x, \cdot)$. If E is only locally compact, choose a compact set $W \subset E$ such that

$$\inf_{x \in E} \inf_{\alpha \in A} \inf_{u \in \mathscr{U}_c(\varepsilon/4)} P_x^{\alpha u}[X_N \in W] \geqslant 1 - \frac{\varepsilon}{8\gamma\|c\|\,k}, \tag{28}$$

and then find a positive integer n such that for $x' \in W$, $\alpha \in A$, $u \in \mathscr{U}_c(\varepsilon/4)$

$$P_{x'}^{\alpha u}\left[\left|\frac{1}{n}\sum_{j=0}^{n-1} \exp\left(\sum_{i=jN+1}^{(j+1)N} \gamma c(X_i, u(X_i))\right) - \int_E E_y^{\alpha u}\left[\exp\left(\sum_{i=1}^{N} \gamma c(X_i, u(X_i))\right)\right]\pi_\alpha^u(dy)\right| \geqslant \frac{\varepsilon}{8}\right] \leqslant \frac{\varepsilon}{8\gamma\|c\|\,k}. \tag{29}$$

4. For the first nN units of time let the control function be $u_1 \in \mathscr{U}_c(\varepsilon/4)$, then use $u_2 \in \mathscr{U}_c(\varepsilon/4)$ for the next nN units of time, and continue by induction until all of the controls in $\mathscr{U}_c(\varepsilon/4)$ are used for nN units of time. Let

$$C_p(u_p) = \sum_{j=0}^{n-1} \exp\left(\sum_{i=((p-1)+j)N+1}^{(p+j)N} \gamma c(X_i, u_p(X_i))\right) \quad \text{for } p = 1, 2, \ldots, k.$$

Let q be such that

$$C_q(u_q) = \min_{p \in \{1, \ldots, k\}} C_p(u_p). \tag{30}$$

5. At time $T = knN$ choose the control function u_q, and for $i \geqslant T$ use the control $\hat{v}_i = u_q(X_i)$.

It is shown that the procedure described by steps 1–5 is feasible and that the control in step 5 is almost optimal. In the case when the state space E is noncompact and (A1), (A3) are satisfied, we shall need the following additional assumption:

(A4) For any $\varepsilon > 0$ there is a compact set $W \subset E$ such that

$$\inf_{x' \in E} \inf_{\alpha \in A} \inf_{u \in E \to U} P_{x'}^{\alpha u}[X_1 \in W] \geqslant 1 - \varepsilon.$$

PROPOSITION 2. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *together with* (A4) *are satisfied, then steps* 1–5 *in the above procedure are feasible.*

Proof. The feasibility of step 1 follows from Theorem 2. By Lemma 1 a positive integer N can be chosen such that $L/N \leqslant \varepsilon/8$ and that step 2 is feasible. By (A2) and (A3) the inequality (6) is satisfied, and consequently (e.g., [8]) for each $u \in \mathcal{U}_c(\varepsilon/4)$ there is a unique invariant measure π_α^u for the transition operator $P^{\alpha u(\cdot)}(x, \cdot)$. Since by (A1) an embedded Markov chain $(X_{iN}, i \in N)$ with transition operator $(P^{\alpha u})^N$ is aperiodic, π_α^u is also the unique invariant measure for $(X_{iN}, i \in N)$.

For $l \in \mathcal{P}(E^N)$ let

$$J_u^\alpha(l) = \sup_{f \in \Phi} \int_{E^N} \log\left(\frac{f(y_1, \ldots, y_N)}{E_{y_N}^{\alpha u}[f(X_1, \ldots, X_N)]}\right) l(dy_1, \ldots, dy_N).$$

By (A1) and the continuity of the control function u, the mapping

$$J_u^{\cdot}(\cdot)\colon A \times \mathcal{P}(E^N) \to R \tag{31}$$

is lower semicontinuous. Moreover, by Lemma 4 the functional $J_\alpha^u(\cdot)$ attains its infimum, that is 0, uniquely at the measure $\hat{l}$ that is given by

$$\hat{l}_\alpha(B) = \int_E E_x^{\alpha u}[1_B(X_1, \ldots, X_N)]\, \pi_\alpha^u(dx) \quad \text{for } B \in \mathcal{B}(E_N).$$

Furthermore, by Lemma 2 the mapping

$$\pi_\cdot^u\colon A \to \mathcal{P}(E^N)$$

is continuous in the variation norm. Therefore, for each $\varphi \in C(E^N)$ the mapping

$$\hat{l}_\cdot(\varphi)\colon A \to R$$

is continuous.

The set of measures

(32)
$$\Gamma_\alpha = \left\{ l \in \mathcal{P}(E^N)\colon \left|\int_E C_N(y_1, \ldots, y_N)\big(l(dy_1, \ldots, dy_n) - \hat{l}(dy_1, \ldots, dy_N)\big)\right| \geqslant \frac{\varepsilon}{8} \right\},$$

where

$$C_N(y_1, \ldots, y_N) = \exp\Big(\sum_{i=1}^N \gamma c(y_i, u(y_i))\Big),$$

is closed in the weak* topology on $\mathscr{P}(E^N)$. If E is compact, then $\mathscr{P}(E^N)$ and Γ_α are compact in the weak* topology and

$$a_u = \inf_{\alpha\in A}\inf_{l\in\Gamma_\alpha} J_u^\alpha(l) > 0. \tag{33}$$

The inequality (33) is satisfied because otherwise by the compactness of A and Γ_α and the lower semicontinuity of the map (31) we have $J_u^{\bar\alpha}(\bar l) = 0$ for some $\bar\alpha\in A$ and $\bar l\in\Gamma_{\bar\alpha}$. By Lemma 3 this implies that $\bar l = \hat l\in\Gamma_{\bar\alpha}$, which contradicts the definition of $\Gamma_{\bar\alpha}$. Proceeding now as in the proof of Theorem 3 of [7], we see that for a $\gamma_u\in(0, a_u)$ there is a positive n_u such that if $n\geqslant n_u$, then

$$\sup_{\alpha\in A}\sup_{x'\in E} P_{x'}^{\alpha u}[S_n^N\in\Gamma_\alpha] \leqslant \exp(-n\gamma_u), \tag{34}$$

where S_n^N is given by (21). Since the family $\mathscr{U}_c(\varepsilon/4)$ is finite, there is a positive integer n such that (29) is satisfied. This completes the proof for E compact.

If E is not compact but only locally compact, then step 3 of the procedure requires additional discussion. Initially it is shown that if (A1) and (A2) are satisfied and if E is locally compact for each $u\in\mathscr{U}_c(\varepsilon/4)$, the assumption (B2) is satisfied. Let $(K_n, n\in N)$ be a sequence of compact sets such that K_n is contained in the interior of K_{n+1} for each $n\in N$ and $\bigcup_{n=1}^\infty K_n = E$ and for a fixed $\bar x\in E$

$$\sup_{\alpha\in A}\sup_{v\in U}\int_{K_n^c} p^{(i)}(\bar x, y, \alpha, v)\eta(dy) \leqslant \frac{1}{(n+1)^3} \quad \text{for } i = 1, \ldots, N,$$

where $p^{(1)}(z, y, \alpha, v) = p(z, y, \alpha, v)$ and

$$p^{(i+1)}(\bar x, y, \alpha, u) = \int_E p^{(i)}(z, y, \alpha, v)\, p(\bar x, z, \alpha, v)\,\eta(dz).$$

Now construct a continuous function $\varphi\colon E\to \boldsymbol{R}$ such that for $y\in K_n\setminus K_{n-1}$

$$n \leqslant \varphi(y) \leqslant n+1,$$

where $K_0 = \emptyset$. Let $\psi\colon E^N\to\boldsymbol{R}$ be given by

$$\psi(y_1, \ldots, y_N) = \varphi(y_1)+\ldots+\varphi(y_N). \tag{35}$$

It follows easily that

$$\sup_{x\in E}\sup_{\alpha\in A} E_x^\alpha[\psi(X_1, \ldots, X_N)] \leqslant \sum_{n=1}^\infty \frac{K+K^2+\ldots+K^N}{(n+1)^3}(n+1) < \infty,$$

so for ψ given by (35) the assumption (B2) is satisfied. By a similar construction one can show that under (A1) and (A4) the assumption (B2) is also satisfied.

The inequality (29) can be obtained as in the compact state space case by using for $u\in\mathscr{U}_c(\varepsilon/4)$ Theorem 4 and following the proof of Theorem 3 in [7]. ∎

Now it is shown that the adaptive procedure of steps 1–5 gives an almost optimal adaptive control.

THEOREM 5. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *together with* (A4) *are satisfied, then the adaptive procedure of steps* 1–5 *defines an* ε*-optimal control in the sense that for* $T = knN$ *the inequality* (2) *is satisfied.*

Proof. For $p \in \{1, \ldots, k\}$ let $B_p \subset \Omega$ be given by

$$B_p = \left\{\left|\frac{1}{n} C_p(u_p) - \int_E E_y^{\alpha^0 u_p}\left[\exp\left(\sum_{i=1}^{N} \gamma c(X_i, u_p(X_i))\right)\right] \pi_{\alpha^0}^{u_p}(dy)\right| < \frac{\varepsilon}{8}\right\}.$$

It is clear that if E is compact, then

$$P_x^{\alpha^0}\left(\bigcap_{p=1}^{k} B_p\right) \geqslant 1 - \frac{\varepsilon}{4\gamma \|c\|}. \tag{36}$$

If E is not compact but only locally compact, then for $p > 1$

$$\begin{aligned} P_x^{\alpha^0}(B_p^c) &= E_x^{\alpha^0}\left[E_x^{\alpha^0}\left[P_{X_{(p-1)nN}}^{\alpha^0}(B_p^c) \mid X_1, \ldots, X_{(p-2)nN}\right]\right] \\ &\leqslant E_x^{\alpha^0}\left[1_W(X_{(p-1)nN}) P_{X_{(p-1)nN}}^{\alpha^0}(B_p^c)\right] + \frac{\varepsilon}{8\gamma \|c\| k} \leqslant \frac{\varepsilon}{4\gamma \|c\|}, \end{aligned}$$

so (36) is satisfied.

For q defined by (30) and $\omega \in \bigcap_{p=1}^{k} B_p$ the following inequality is satisfied for $p = \{1, \ldots, k\}$:

$$\begin{aligned} &\int_E E_y^{\alpha^0 u_q}\left[\exp\left(\sum_{i=1}^{N} \gamma c(X_i, u_q(X_i))\right)\right] \pi_{\alpha^0}^{u_q}(dy) \\ &\qquad \leqslant \frac{\varepsilon}{4} + \int_E E_y^{\alpha^0 u_p}\left[\exp\left(\sum_{i=1}^{N} \gamma c(X_i, u_p(X_i))\right)\right] \pi_{\alpha^0}^{u_p}(dy). \end{aligned}$$

From (26) it follows that

$$\lambda_{\alpha^0}^{u_q} \leqslant \varepsilon/4 + \varepsilon/4 + \lambda_{\alpha^0}^{u_p} = \tfrac{1}{2}\varepsilon + \lambda_{\alpha^0}^{u_p} \quad \text{for } p \in \{1, \ldots, k\} \text{ and } \omega \in \bigcap_{p=1}^{k} B_p.$$

Thus

$$\begin{aligned} &\limsup_{n\to\infty} E_x^{\alpha^0}\left[\frac{1}{n}\log E_x^{\alpha^0}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, \hat{v}_i)\right)\right] \mid X_1, \ldots, X_T\right] \\ &\leqslant \limsup_{n\to\infty} E_x^{\alpha^0}\left[1_{\bigcap_{p=1}^{k} B_p} \frac{1}{n}\log E_{X_T}^{\alpha^0}\left[\exp\left(\sum_{i=0}^{n-1} \gamma c(X_i, \hat{v}_{i+T})\right)\right]\right] \\ &\quad + E_x^{\alpha^0}\left[1_{\bigcup_{p=1}^{k} B_p^c} \frac{1}{n}\log\exp(n\gamma \|c\|)\right] \\ &\leqslant E_x^{\alpha^0}\left[1_{\bigcap_{p=1}^{k} B_p} \lambda_{\alpha^0}^{u_q}\right] + \gamma \|c\| P_x^{\alpha^0}\left(\bigcup_{p=1}^{k} B_p^c\right) \leqslant \frac{1}{2}\varepsilon + \inf_{p\in\{1,\ldots,k\}} \lambda_{\alpha^0}^{u_p} + \frac{\varepsilon}{4} \leqslant \lambda_{\alpha^0} + \varepsilon. \quad \blacksquare \end{aligned}$$

4. ADAPTIVE CONTROL WITH ESTIMATION

Let $u: E \to U$ be continuous and $\alpha, \alpha' \in A$ and let $K: U \times A \times A \to R$ be given by

$$(37)\qquad K^u(\alpha, \alpha') := \int_E \int_E \log p\big(x, y, \alpha', u(x)\big)\, p\big(x, y, \alpha, u(x)\big)\, \eta(dy)\, \pi_\alpha^u(dx).$$

The function $K^{\cdot}(\cdot, \cdot)$ has the following continuity property:

LEMMA 5. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied and* $u: E \to U$ *is continuous, then*

$$K^u: A \times A \to R$$

is continuous. Furthermore, if $K^u(\alpha, \alpha) - K^u(\alpha, \alpha') = 0$, *then for* $x \in \operatorname{supp}(\eta)$,

$$P^{\alpha u}(x, \cdot) = P^{\alpha' u}(x, \cdot).$$

Proof. The continuity of $K^u(\cdot, \cdot)$ follows from (A1), (A2) and Lemma 2. To verify the second claim in the lemma, we infer from Jensen's inequality that

$$\int_E \log\left(\frac{p\big(x, y, \alpha, u(x)\big)}{p\big(x, y, \alpha', u(x)\big)}\right) p\big(x, y, \alpha, u(x)\big)\, \eta(dy) \geqslant 0.$$

If $K^u(\alpha, \alpha) - K^u(\alpha, \alpha') = 0$, then

$$(38)\qquad \int_E \log\left(\frac{p\big(x, y, \alpha, u(x)\big)}{p\big(x, y, \alpha', u(x)\big)}\right) p\big(x, y, \alpha, u(x)\big)\, \eta(dy) = 0$$

for π_α^u almost all x, and thus for η almost all x and, by the continuity of the transition densities, for all $x \in \operatorname{supp}(\eta)$. Since $\log(\cdot)$ is a strictly convex function, (38) is satisfied only if

$$p\big(x, y, \alpha, u(x)\big) = p\big(x, y, \alpha', u(x)\big) \quad \text{for all } x, y \in \operatorname{supp}(\eta).$$

Therefore, $P^{\alpha u}(x, \cdot) = P^{\alpha' u}(x, \cdot)$ for all $x \in \operatorname{supp}(\eta)$. ∎

The estimation procedure that is used subsequently is based on the following lemma:

LEMMA 6. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *are satisfied, then for each continuous* $u \in E \to R$ *and* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that*

$$(39)\qquad \text{if } K^u(\alpha, \alpha) - K^u(\alpha, \alpha') < \delta \text{ for some } \alpha, \alpha' \in A, \text{ then } |\lambda_\alpha^u - \lambda_{\alpha'}^u| < \varepsilon.$$

Proof. Assume that there is an $\varepsilon > 0$ such that for no $\delta > 0$ the relation (39) is satisfied, so there are two sequences $(\alpha_n, n \in N)$ and $(\alpha_n', n \in N)$ that converge to α and α', respectively, such that

$$\lim_{n \to \infty} \big(K^u(\alpha_n, \alpha_n) - K^u(\alpha_n, \alpha_n')\big) = 0$$

and $|\lambda^u_{\alpha_n} - \lambda^u_{\alpha'_n}| \geqslant \varepsilon$ for all n. By the continuity of K^u from Lemma 5 we obtain $K^u(\alpha, \alpha) - K^u(\alpha, \alpha') = 0$ so that $P^{\alpha u}(x, \cdot) = P^{\alpha' u}(x, \cdot)$ for $x \in \operatorname{supp}(\eta)$, and consequently $\lambda^u_\alpha = \lambda^u_{\alpha'}$. By Corollary 3 there is a contradiction. ∎

Now an adaptive procedure with estimation is described. Given $\varepsilon > 0$ and an initial state $x \in E$, the following steps are used:

1. Determine the finite family $\mathscr{U}_c(\varepsilon/4) = \{u_1, \ldots, u_k\}$ of continuous $(\varepsilon/4)$-optimal control functions by Theorem 2.

2. Determine a $\delta > 0$ such that for $u \in \mathscr{U}_c(\varepsilon/4)$ and $\alpha, \alpha' \in A$

$$\text{if } K^u(\alpha, \alpha) - K^u(\alpha, \alpha') \leqslant \delta, \text{ then } |\lambda^u_\alpha - \lambda^u_{\alpha'}| \leqslant \varepsilon/8. \tag{40}$$

3. Determine a finite family $A_f = \{\alpha_1, \ldots, \alpha_{k'}\} \subset A$ and a finite cover A_i, $i = 1, \ldots, k'$, of A such that $\alpha_i \in A_i$ for $i \in \{1, \ldots, k'\}$ and for $\alpha \in A$ there is a $j \in \{1, \ldots, k'\}$ with $\alpha \in A_j$ and for each $u \in \mathscr{U}_c(\varepsilon/4)$

$$K^u(\alpha, \alpha) - K^u(\alpha, \alpha_j) \leqslant \delta/2 \tag{41}$$

and

$$\sup_{\bar{\alpha} \in A} |K^u(\alpha_j, \bar{\alpha}) - K^u(\alpha, \bar{\alpha})| \leqslant \delta/16. \tag{42}$$

4. For $u \in \mathscr{U}(\varepsilon/4)$ define the estimator $\hat{\alpha}^u$ as

$$\hat{\alpha}^u_n = \Big\{\alpha_j \in A_f \colon \prod_{i=0}^{n-1} p\big(X_i, X_{i+1}, u(X_i), \alpha_j\big) = \max_{p \in \{1, \ldots, k\}} \prod_{i=0}^{n-1} p\big(X_i, X_{i+1}, u(X_i), \alpha_p\big)\Big\}.$$

If E is compact, then find a positive integer N such that for $u \in \mathscr{U}_c(\varepsilon/4)$

$$\sup_{\alpha \in A} \sup_{x \in E} P^{\alpha u}_\alpha \left(|\lambda^u_\alpha - \lambda^u_{\hat{\alpha}^u_N}| \geqslant \frac{\varepsilon}{8} \right) \leqslant \frac{\varepsilon}{2\,\|c\|\,k}. \tag{43}$$

If E is only locally compact, choose a compact set $W \subset E$ such that

$$\inf_{x' \in E} \inf_{\alpha \in A} \inf_{u \in \mathscr{U}_c(\varepsilon/4)} P^{\alpha u}_{x'} \{X_1 \in W\} \geqslant 1 - \frac{\varepsilon}{4\,\|c\|\,k}. \tag{44}$$

Then choose an $N \in N$ such that for $u \in \mathscr{U}_c(\varepsilon/4)$

$$\sup_{\alpha \in A} \sup_{x \in W} P^{\alpha u}_x \left(|\lambda^u_\alpha - \lambda^u_{\hat{\alpha}^u_N}| \geqslant \frac{\varepsilon}{8} \right) \leqslant \frac{\varepsilon}{4\,\|c\|\,k}. \tag{45}$$

5. Use each of the control functions $u \in \mathscr{U}_c(\varepsilon/4)$ for N successive units of time. For $p = 1, \ldots, k$ define

$$\hat{\alpha}_p = \Big\{\alpha \in A_f \colon \prod_{i=N(q-1)}^{N_q - 1} p\big(X_i, X_{i+1}, u_p(X_i), \alpha\big) = \max_{\bar{\alpha} \in A_f} \prod_{i=N(q-1)}^{N_q - 1} p\big(X_i, X_{i+1}, u_p(X_i), \bar{\alpha}\big)\Big\},$$

where ties are settled by choosing $\alpha \in A_f$ with the smallest index. Choose $q \in \{1, 2, \ldots, k\}$ such that

$$\lambda_{\bar{\alpha}_q}^{u_q} = \min_{p \in \{1,2,\ldots,k\}} \lambda_{\alpha_p}^{u_p}.$$

6. At time $T = kN$ choose the control function u_q and for $i > kN$ use the control $\hat{v}_i = u_q(X_i)$.

It is shown that this adaptive procedure is feasible.

PROPOSITION 3. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *together with* (A4) *are satisfied, then steps* 1–6 *in the above adaptive estimation and the control procedure are feasible.*

Proof. Step 1 follows from Theorem 2. Step 2 can be performed by Lemma 6. Step 3 is feasible by Lemma 5 and the compactness of A. Step 4 is feasible from the proof of Theorem 10 of [7] because, as in the proof of Proposition 2, it follows that if (A1) and (A2) or (A1), (A3) and (A4) are satisfied, then (B4) of [7] is also satisfied, so a large deviations upper bound for empirical distributions of pairs of consecutive states (Theorems 5 and 6 of [7]) can be applied. Thus, there is a $\gamma > 0$ and an $N \in N$ such that for $n \geqslant N$, $i, j \in \{1, 2, \ldots, k'\}$, and $u \in \mathcal{U}_c(\varepsilon/4)$

$$\sup_{\alpha \in A_i} \sup_{x \in W} P_x^{\alpha u}\left(\left|\frac{1}{n} \sum_{m=0}^{n-1} \log p\big(X_m, X_{m+1}, u(X_m), \alpha_j\big) - K^u(\alpha_i, \alpha_j)\right| \geqslant \frac{\delta}{8}\right) \leqslant e^{-n\gamma}.$$

Using the methods of proof of Theorem 10 in [7] it follows that for $u \in \mathcal{U}_c(\varepsilon/4)$

$$\sup_{\alpha \in A_i} \sup_{x \in W} P_x^{\alpha u}\big(K^u(\alpha, \alpha) - K^u(\alpha, \hat{\alpha}_n^u) > \delta\big) \leqslant k' e^{-n\gamma}.$$

Using this inequality and (40) we infer that, for N sufficiently large, (43) and (45) are satisfied. The existence of a compact set W in (44) follows directly from (A2) or (A4). ∎

The almost optimality of the adaptive procedure with steps 1–6 is given now.

THEOREM 6. *If* (A1) *and* (A2) *or* (A1) *and* (A3) *together with* (A4) *are satisfied, then the adaptive procedure with estimation given by steps* 1–6 *is* ε*-optimal, that is, for* $T = kN$ *the inequality* (2) *is satisfied.*

Proof. Let $B_p \subset \Omega$ be given by

$$B_p = \{|\lambda_{\alpha^0}^{u_p} - \lambda_{\bar{\alpha}_p}^{u_p}| \leqslant \varepsilon/8\}.$$

Using the methods of proof of Theorem 5 and (43)–(45) we obtain

$$P_x^{\alpha^0}\Big(\bigcap_{p=1}^{k} B_k\Big) \geqslant 1 - \frac{\varepsilon}{4\gamma \|c\|}.$$

If $\omega \in \bigcap_{p=1}^{k} B_k$, then

$$\lambda_{\alpha^0}^{u_q} \leqslant \lambda_{\bar{\alpha}_q}^{u_q} + \varepsilon/8 = \inf_{p \in \{1,\ldots,k\}} \lambda_{\bar{\alpha}_p}^{u_p} + \varepsilon/8 \leqslant \inf_{p \in \{1,\ldots,k\}} \lambda_{\alpha^0}^{u_p} + \varepsilon/4 \leqslant \lambda_\alpha + \varepsilon/2.$$

Thus

$$\limsup_{n \to \infty} E_x^{\alpha^0} \left[\frac{1}{n} \log E^{\alpha^0} \left[\exp \left(\sum_{i=0}^{n-1} \gamma c(X_i, \hat{v}_i) \right) \mid X_1, \ldots, X_T \right] \right]$$
$$\leqslant E_x^{\alpha^0} [1_{\bigcap_{p=1}^{k} B_p} \lambda_{\alpha^0}^{u_q}] + \gamma \|c\| P_x^{\alpha^0} \left(\bigcup_{p=1}^{k} B_p^c \right) \leqslant \lambda_\alpha + \varepsilon. \blacksquare$$

REFERENCES

[1] G. B. Di Masi and Ł. Stettner, *Bayesian ergodic adaptive control of discrete time Markov processes*, Stochastics Stochastic Rep. 54 (1995), pp. 301–316.

[2] G. B. Di Masi and Ł. Stettner, *Risk sensitive control of discrete time Markov processes with infinite horizon*, SIAM J. Control Optim. 38 (2000), pp. 61–78.

[3] G. B. Di Masi and Ł. Stettner, *Risk sensitive control of discrete time Markov processes with small risk*, Systems Control Lett. 40 (2000), pp. 15–20.

[4] M. D. Donsker and S. R. S. Varadan, *Asymptotic evaluation of certain Markov process expectations for large time. I*, Comm. Pure Appl. Math. 28 (1975), pp. 1–47.

[5] M. D. Donsker and S. R. S. Varadan, *Asymptotic evaluation of certain Markov process expectations for large time. III*, Comm. Pure Appl. Math. 28 (1976), pp. 384–461.

[6] T. E. Duncan, B. Pasik-Duncan and Ł. Stettner, *Discretized maximum likelihood and almost optimal adaptive control of ergodic Markov models*, SIAM J. Control Optim. 36 (1998), pp. 422–446.

[7] T. E. Duncan, B. Pasik-Duncan and Ł. Stettner, *Adaptive control of discrete time Markov processes by the large deviation method*, Appl. Math. 27 (3) (2000), pp. 265–285.

[8] O. Hernandez-Lerma, *Adaptive Control Processes*, Springer-Verlag, New York–Berlin 1989.

T. E. Duncan and B. Pasik-Duncan
Department of Mathematics
University of Kansas
Lawrence, KS 66045, U.S.A.

Ł. Stettner
Institute of Mathematics
Polish Academy of Sciences
Śniadeckich 8
00-950 Warsaw, Poland

Received on 6.1.2001;
revised version on 31.7.2001