J. Adolfo Minjárez-Sosa

Approximation and estimation in Markov control processes under a discounted criterion

Kybernetika, Vol. 40 (2004), No. 6, [681]--690

Persistent URL: http://dml.cz/dmlcz/135626

KYBERNETIKA — VOLUME 40 (2004), NUMBER 6, PAGES 681 - 690

APPROXIMATION AND ESTIMATION IN MARKOV CONTROL PROCESSES UNDER A DISCOUNTED CRITERION*

J. ADOLFO MINJÁREZ-SOSA

We consider a class of discrete-time Markov control processes with Borel state and action spaces, and $\Re^k$-valued i.i.d. disturbances with unknown density ρ. Supposing possibly unbounded costs, we combine suitable density estimation methods of ρ with approximation procedures of the optimal cost function, to show the existence of a sequence $\{\hat{f}_t\}$ of minimizers converging to an optimal stationary policy f_∞.

Keywords: Markov control processes, density estimation, discounted cost criterion

AMS Subject Classification: 93E10, 90C40

1. INTRODUCTION

To study a stochastic control problem associated to a Markov control model under discounted cost criterion, typically it is required the following: First to prove that the optimal cost function V^* is a solution to the optimality equation (Dynamic Programming Equation) – problem 1; and then to solve a minimization problem to calculate optimal policies – problem 2.

However, the solution of problems 1 and 2 generally is difficult, and it is therefore of great importance to propose efficient approximation algorithms for V^* and construction methods of optimal policies.

Our main objective in this paper is to study both problems for a class of discrete-time Markov control processes of the form

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, \ldots, \tag{1}$$

where F is a known function, x_t, a_t and ξ_t are the state, action and random disturbance at time t, respectively. Moreover, $\{\xi_t\}$ is an observable sequence of independent and identically distributed (i.i.d.) random vectors in $\Re^k$ having density ρ which is *unknown* to the controller. In addition, we suppose that the one-stage cost (and therefore the optimal cost V^*) is unbounded. In this context, our approach consists in the following. First, we introduce an approximation algorithm of V^* based in

*Work supported partially by Consejo Nacional de Ciencia y Tecnologia (CONACyT) under Grant 37239E.

the combination of suitable density estimation methods of ρ with a value iteration scheme. Then, this approximation algorithm is used to show the existence of a sequence of minimizers $\{\hat{f}_t\}$ (which depends of the estimators ρ_t of ρ) converging, in the sense of Schäl [13], to an optimal stationary policy f_∞.

The assumption of unbounded costs generates serious difficulties. For instance, the nice contractive-operator techniques do not work for the discounted criterion. For this reason, in previous papers where similar problems are analyzed (see, e. g., [5, 12]), it was necessary to impose restrictive conditions on the unknown density ρ and apply a density estimation process which is difficult to implement. This set of assumptions might be strong even for very simple applied problems. In contrast, our results here are obtained exploiting some easy facts in the theory of density estimation. Others papers where similar problems are studied but considering bounded costs are, for instance, [1, 4, 7, 8, 11].

The paper is organized as follows. In Section 2 we introduce the Markov control model we deal with. Next, Section 3 contains the assumptions on the control model and some preliminary results on the discounted criterion, which are used to state our main results in Section 4. The proofs are presented in Section 5, and finally, an example of a storage system is introduced in Section 6 to illustrate our results.

2. MARKOV CONTROL MODELS

Notation. Given a Borel space X (that is, a Borel subset of a complete and separable metric space) its Borel sigma-algebra is denoted by $\mathcal{B}(X)$, and "measurable", for either sets or functions, means "Borel measurable". In addition, we denote by $\mathbb{B}(X)$ the space of real-valued bounded measurable functions on X with the supremum norm $\|v\| := \sup_x |v(x)|$.

Control model. Let

$$\mathcal{M} := \left(X, A, \{A(x) \subset A \mid x \in X\}, \Re^k, F, \rho, c\right) \tag{2}$$

be a discrete-time Markov control model where the state space X and the action or control space A are Borel spaces endowed with their Borel σ-algebras. To each $x \in X$ it is associated a nonempty set $A(x) \in \mathcal{B}(A)$ whose elements are the admissible controls when the system is in state x. The set

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

of admissible state-action pairs is assumed to be a Borel subset of the Cartesian product of X and A. The dynamics is defined by the system equation (1) where $F : X \times A \times \Re^k \to X$ is a given (known) measurable function, and $\{\xi_t\}$ is a sequence of independent and identically distributed (i.i.d.) random vectors (r.v.'s) on a probability space $(\Omega, \mathcal{F}, P)$, with values in $\Re^k$ and common distribution with a *unknown* density ρ. Finally, the cost-per-stage $c(x, a)$ is a nonnegative measurable real-valued function on $\mathbb{K}$, possibly unbounded.

Throughout the paper, the probability space $(\Omega, \mathcal{F}, P)$ is fixed and *a. s.* means *almost surely with respect to* P. In addition, we assume that the realizations $\xi_0, \xi_1, \ldots$ of the disturbance process and the states $x_0, x_1, \ldots$ are completely observable.

Control policies. We define the spaces of admissible histories up to time t by $\mathbb{H}_0 := X$ and $\mathbb{H}_t := (\mathbb{K} \times \Re^k)^t \times X$, for $t \in \mathbb{N} := \{1, 2, \dots\}$. A typical element of $\mathbb{H}_t$ is written as $h_t = (x_0, a_0, \xi_0, \dots, x_{t-1}, a_{t-1}, \xi_{t-1}, x_t)$. A *control policy* $\pi = \{\pi_t\}$ is a sequence of measurable functions $\pi_t : \mathbb{H}_t \to A$ such that $\pi_t(h_t) \in A(x_t)$, for all $h_t \in \mathbb{H}_t$, $t \in \mathbb{N}$. We denote by Π the set of all control policies.

Let $\mathbb{F}$ be the family of measurable functions $f : X \to A$ such that $f(x) \in A(x)$ for all $x \in X$. A sequence $\{f_t\}$ of functions $f_t \in \mathbb{F}$ is called a *Markov policy.* A Markov policy $\{f_t\}$ is said to be *stationary* if $f_t = f$ for all $t = 0, 1, \dots$ and some $f \in \mathbb{F}$. In this case we use the notation

$$c(x, f_t) := c(x, f_t(x)) \quad \text{and} \quad F(x, f_t, s) := F(x, f_t(x), s)$$

for all $x \in X$, $s \in \Re^k$, and $t \geq 0$.

3. DISCOUNTED OPTIMALITY CRITERION

When using a policy $\pi \in \Pi$, given the initial state $x_0 = x$, we define the total expected α-discounted cost as

$$V(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right], \tag{3}$$

where $\alpha \in (0, 1)$ is the so-called discount factor, and E_x^π denotes the expectation operator with respect to the probability measure P_x^π induced by the policy π, given the initial state $x_0 = x$ (see, e.g., [3] for the construction of P_x^π).

The optimal control problem associated to the control model $\mathcal{M}$, is then to find an optimal policy $\pi^* \in \Pi$ such that $V(\pi^*, x) = V^*(x)$ for all $x \in X$, where

$$V^*(x) := \inf_{\pi \in \Pi} V(\pi, x), \quad x \in X,$$

is the optimal α-discounted cost, which we call *value function.*

Assumptions. To guarantee the existence of "measurable minimizers", we need the following standard continuity and compactness conditions on the components of the control model $\mathcal{M}$.

Assumption 3.1. (a) For every $x \in X$, the one-stage cost $c(x, a)$ is nonnegative and continuous on $a \in A(x)$. Moreover, there exist a measurable function $W : X \to [1, \infty)$ and constants $\tilde{c} > 0$ and $\beta > 0$, such that $0 < \alpha\beta < 1$, $\sup_{A(x)} c(x, a) \leq \tilde{c} W(x)$ and

$$\int_{\Re^k} W[F(x, a, s)] \rho(s)\, ds \leq \beta W(x).$$

(b) For each $x \in X$, $A(x)$ is a compact set.

(c) For each $x \in X$ and $v \in \mathbb{B}(X)$, the function $a \to \int_{\Re^k} v[F(x, a, s)] \rho(s)\, ds$ is continuous and bounded on $A(x)$.

(d) For each $x \in X$, the function $a \to \int_{\Re^k} W[F(x,a,s)]\rho(s)\,ds$ is continuous on $A(x)$.

We denote by $\mathbb{B}_W(X)$ the normed linear space of all measurable function $u : X \to \Re$ with

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

A first consequence of Assumption 3.1 is the following (see, e. g., [10]):

Proposition 3.2. Suppose that Assumption 3.1 holds. Then:
a) $V^* \in \mathbb{B}_W(X)$ is a solution to the α-discounted optimality equation

$$V^*(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_{\Re^k} V^*(F(x,a,s))\rho(s)\,ds \right\}, \quad x \in X. \tag{4}$$

b) There exists $f \in \mathbb{F}$ such that $f(x) \in A(x)$ attains the minimum in (4), i. e.,

$$V^*(x) = c(x,f) + \alpha \int_{\Re^k} V^*(F(x,f,s))\rho(s)\,ds, \quad x \in X, \tag{5}$$

and moreover, the stationary policy $\{f\}$ is optimal.

4. APPROXIMATION AND ESTIMATION

Remark 4.1. In [9] there were presented several approximation schemes to the value function V^*, for instance, the "*recursive bounded-cost approximations*" defined as follows. Let $\{\bar{c}_n\}_{n \in \mathbb{N}}$ be a sequence of nonnegative bounded and continuous functions on $\mathbb{K}$ such that $\bar{c}_n \nearrow c$. We define the sequence $\{u_n\}$ of functions on $\mathbb{B}(X)$ as:

$$u_0 \equiv 0;$$

$$u_n(x) := \min_{a \in A(x)} \left\{ \bar{c}_n(x,a) + \alpha \int_{\Re^k} u_{n-1}(F(x,a,s))\rho(s)\,ds \right\}, \quad x \in X,\ n \geq 1. \tag{6}$$

Then, under Assumption 3.1, $u_n \nearrow V^*$. Our approach is motived by this approximation scheme.

Let $\xi_0, \xi_1, \ldots, \xi_{n-1}$ be independent r.v.'s (observed up to time $n-1$) with the unknown density ρ. We consider the control model

$$\mathcal{M}_n = \left(X, A, \{A(x) \subset A \,|\, x \in X\}, \Re^k, F, \rho_n, c_n\right)$$

satisfying the following conditions. The state space X, the control space A and the function F are as in (2); $\rho_n(s) := \rho_n(s; \xi_0, \xi_1, \ldots, \xi_{n-1})$, $s \in \Re^k$, is an estimator of ρ such that, for some $\gamma > 0$,

$$E \int_{\Re^k} |\rho_n(s) - \rho(s)|\,ds = O(n^{-\gamma}) \quad \text{as} \quad n \to \infty; \tag{7}$$

and, finally, $c_n : \mathbb{K} \to \Re$ is the truncated cost defined as

$$c_n(x,a) := \min\{c(x,a), n\}, \quad (x,a) \in \mathbb{K}. \tag{8}$$

Estimators satisfying (7) are given, for instance, in [2, 6].

Remark 4.2. a) In particular, observe that (see Remark 4.1)

$$U_n \nearrow V^* \quad \text{as} \quad n \to \infty, \tag{9}$$

where $\{U_n\}$ is the sequence of function on $\mathbb{B}(X)$ defined in (6) corresponding to the truncated cost (8). That is,

$$U_0 \equiv 0;$$
$$U_n(x) := \min_{a \in A(x)} \left\{ c_n(x,a) + \alpha \int_{\Re^k} U_{n-1}(F(x,a,s))\rho(s)\,\mathrm{d}s \right\}, \quad x \in X,\ n \geq 1. \tag{10}$$

In fact, since $c_n(x,a) \leq n$ for each $n \geq 0$, it is easy to see that

$$U_n(x) \leq \sum_{k=1}^{n} k\alpha^{n-k} \leq \sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \quad x \in X. \tag{11}$$

b) In addition, under Assumption 3.1, $\sup_{A(x)} c_n(x,a) \leq \bar{c}W(x)$ for all $x \in X$, $n \geq 0$. Furthermore, $\{U_n\}$ is a sequence of functions on $\mathbb{B}_W(X)$, such that,

$$U_n(x) \leq \frac{\bar{c}W(x)}{1-\alpha\beta}.$$

For each fixed $t \geq 0$, we define the sequence $\{V_n^{\rho_t}\}$ of functions on $\mathbb{B}(X)$ as:

$$V_0^{\rho_t} \equiv 0;$$
$$V_n^{\rho_t}(x) := \min_{a \in A(x)} \left\{ c_n(x,a) + \alpha \int_{\Re^k} V_{n-1}^{\rho_t}(F(x,a,s))\rho_t(s)\,\mathrm{d}s \right\}, \quad x \in X,\ n \geq 1. \tag{12}$$

Now, let us choose an arbitrary real number $\nu \in (0, \gamma/3)$ (γ as in (7)) and define a sequence $\{n_t\}$ of integer numbers as $n_t := [t^\nu]$, where $[x]$ represents the integer part of x.

Remark 4.3. Applying standard arguments on the existence of minimizers (see, e.g., [7, 9] and references therein), under Assumption 3.1 we have that for each $t \in \mathbb{N}$, there exists $\hat{f}_t \equiv f_{n_t}^{\rho_t} \in \mathbb{F}$ such that

$$V_{n_t}^{\rho_t}(x) = c_{n_t}(x, \hat{f}_t) + \alpha \int_{\Re^k} V_{n_t-1}^{\rho_t}(F(x, \hat{f}_t, s))\rho_t(s)\,\mathrm{d}s, \quad \forall x \in X, \tag{13}$$

where the minimization is done for every $\omega \in \Omega$. Moreover, by a result of Schäl [13], there is a stationary policy $\{f_\infty\}$ for the control model $\mathcal{M}$ such that for each $x \in X$, $f_\infty(x) \in A(x)$ is an accumulation point of $\left\{\hat{f}_t(x)\right\}$. That is, there exists a subsequence $\{t_i\}$ of $\{t\}$ ($t_i = t_i(x)$) such that $\hat{f}_{t_i}(x) \to f_\infty(x)$ as $i \to \infty$.

Theorem 4.4. Suppose that Assumption 3.1 holds. Then:
a) $E\left\|V_{n_t}^{\rho_t} - U_{n_t}\right\| \to 0$ as $t \to \infty$.
b) For each $x \in X$,

$$E\left|V_{n_t}^{\rho_t}(x) - V^*(x)\right| \to 0 \quad \text{as} \quad t \to \infty.$$

c) If moreover, the set-valued mapping $x \longmapsto A(x)$ is upper semicontinuous and F is continuous in $a \in A(x)$ for all $x \in X$, then the stationary policy $\{f_\infty\}$ is optimal for the model $\mathcal{M}$.

Remark 4.5. (a) Observe that from (13), letting $t_i = i$ for notational convenience, we have for each $i \in \mathbb{N}$,

$$V_{n_i}^{\rho_i}(x) = c_{n_i}(x, \hat{f}_i) + \alpha \int_{\Re^k} V_{n_i-1}^{\rho_i}(F(x, \hat{f}_i, s))\rho_i(s)\, ds \text{ a.s.}, \quad \forall i \geq 0,\ x \in X. \quad (14)$$

(b) Upper semi-continuity of $x \longmapsto A(x)$ means: for each open set $A' \subset A$, the set $\{x \in X : A(x) \subset A'\}$ is open in X. This assumption together Assumption 3.1 implies that the value function V^* is lower semi continuous (see, e. g., [9]).

5. PROOF OF THEOREM 4.4

a) From (12) and (10), adding and subtracting the term $\alpha \int_{\Re^k} U_{n_t-1}(F(x,a,s))\rho_t(s)\, ds$ we have

$$\begin{aligned}
\left|V_{n_t}^{\rho_t}(x) - U_{n_t}(x)\right| &\leq \sup_{a \in A(x)} \left| \int_{\Re^k} V_{n_t-1}^{\rho_t}(F(x,a,s))\rho_t(s)\, ds - \int_{\Re^k} U_{n_t-1}(F(x,a,s))\rho(s)\, ds \right| \\
&\leq \sup_{a \in A(x)} \left\{ \int_{\Re^k} \left|V_{n_t-1}^{\rho_t}(F(x,a,s)) - U_{n_t-1}(F(x,a,s))\right| \rho_t(s)\, ds \right. \\
&\quad \left. + \int_{\Re^k} U_{n_t-1}(F(x,a,s))\left|\rho_t(s) - \rho(s)\right|\, ds \right\} \\
&\leq \left\|V_{n_t-1}^{\rho_t} - U_{n_t-1}\right\| + \left\|U_{n_t-1}\right\| \int_{\Re^k} \left|\rho_t(s) - \rho(s)\right|\, ds, \quad t \geq 0.
\end{aligned}$$

Hence,

$$\left\|V_{n_t}^{\rho_t} - U_{n_t}\right\| \leq \left\|V_{n_t-1}^{\rho_t} - U_{n_t-1}\right\| + \left\|U_{n_t-1}\right\| \int_{\Re^k} \left|\rho_t(s) - \rho(s)\right|\, ds, \quad t \geq 0.$$

Iterating this inequality and using that $V_0^{\rho_t} = U_0 = 0$, we obtain

$$\left\|V_{n_t}^{\rho_t} - U_{n_t}\right\| \leq (\|U_0\| + \cdots + \|U_{n_t-1}\|) \int_{\Re^k} \left|\rho_t(s) - \rho(s)\right|\, ds, \quad t \geq 0,$$

which in turn yields

$$\begin{aligned}
\left\|V_{n_t}^{\rho_t} - U_{n_t}\right\| &\leq n_t \left\|U_{n_t-1}\right\| \int_{\Re^k} \left|\rho_t(s) - \rho(s)\right|\, ds \\
&\leq \frac{n_t^2(n_t-1)}{2} \int_{\Re^k} \left|\rho_t(s) - \rho(s)\right|\, ds \text{ a.s.} \quad t \geq 0, \qquad (15)
\end{aligned}$$

since $\{U_n\}$ is a increasing sequence, and from (11).

Now, by the definition of n_t we have

$$\frac{n_t^2(n_t-1)}{2} = O(t^{3\nu}) \quad \text{as} \quad t \to \infty.$$

Thus, from (7) and taking expectation on both sides of (15), we get

$$E\left\|V_{n_t}^{\rho_t} - U_{n_t}\right\| = O(t^{3\nu})O(t^{-\gamma}) = O(t^{3\nu-\gamma}) \to 0 \quad \text{as} \quad t \to \infty$$

because $\nu < \gamma/3$ (see the definition of n_t). This completes the proof of the part (a).

b) This result is a consequence of part (a) and (9). Indeed, for each $x \in X$ and $t \geq 0$,

$$\left|V_{n_t}^{\rho_t}(x) - V^*(x)\right| \leq \left|V_{n_t}^{\rho_t}(x) - U_{n_t}(x)\right| + \left|U_{n_t}(x) - V^*(x)\right| \quad \text{a.s.}$$

Taking expectation on both sides of this inequality and letting $t \to \infty$, we obtain the desired result.

c) We fix an arbitrary $x \in X$. Adding and subtracting the terms $\int_{\Re^k} U_{n_i-1}(F(x,a,s))\rho(s)\mathrm{d}s$ and $\int_{\Re^k} U_{n_i-1}(F(x,a,s))\rho_i(s)\mathrm{d}s$, we have, for each $i \geq 0$,

$$\begin{aligned}
&\left|\int_{\Re^k} V^*(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s - \int_{\Re^k} V_{n_i-1}^{\rho_i}(F(x,\hat{f}_i,s))\rho_i(s)\,\mathrm{d}s\right| \\
&\leq \left|\int_{\Re^k} V^*(F(x,\hat{f}_i(x),s))\rho(s)\,\mathrm{d}s - \int_{\Re^k} U_{n_i-1}(F(x,\hat{f}_i(x),s))\rho(s)\,\mathrm{d}s\right| \\
&\qquad + \sup_{a\in A(x)} \int_{\Re^k} U_{n_i-1}(F(x,a,s))\left|\rho_i(s) - \rho(s)\right|\,\mathrm{d}s \\
&\qquad + \sup_{a\in A(x)} \int_{\Re^k} \left|U_{n_i-1}(F(x,a,s)) - V_{n_i-1}^{\rho_i}(F(x,a,s))\right|\rho_i(s)\,\mathrm{d}s. \qquad (16)
\end{aligned}$$

Now, the facts $\hat{f}_i \to f_\infty$, $U_n \in \mathbb{B}_W(X)$ (see Remark 4.2(b)), Fatou's Lemma (see Lemma 8.3.7 in [10]) and (9) yield,

$$\left|\int_{\Re^k} V^*(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s - \int_{\Re^k} U_{n_i-1}(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s\right| \to 0 \quad \text{as} \quad i \to \infty. \quad (17)$$

Now, from (11) and (15) we get

$$\begin{aligned}
\beta_i &:= \sup_{a\in A(x)} \int_{\Re^k} U_{n_i-1}(F(x,a,s))\left|\rho_i(s) - \rho(s)\right|\,\mathrm{d}s \\
&\leq \frac{(n_i-1)n_i}{2}\int_{\Re^k} \left|\rho_i(s) - \rho(s)\right|\,\mathrm{d}s \leq n_i^2 \int_{\Re^k} \left|\rho_i(s) - \rho(s)\right|\,\mathrm{d}s, \quad x \in X; \qquad (18)
\end{aligned}$$

and

$$\delta_i := \sup_{a\in A(x)} \int_{\Re^k} \left|U_{n_i-1}(F(x,a,s)) - V^{\rho_i}_{n_i-1}(F(x,a,s))\right| \rho_i(s)\,\mathrm{d}s$$
$$\leq \frac{(n_i-1)^2(n_i-2)}{2} \int_{\Re^k} |\rho_i(s)-\rho(s)|\,\mathrm{d}s \leq n_i^3 \int_{\Re^k} |\rho_i(s)-\rho(s)|\,\mathrm{d}s, \quad x\in X. \tag{19}$$

Thus, taking expectation on both sides of (18) and (19), the definition of n_i together with (7) implies,

$$E\beta_i = O(i^{2\nu})O(i^{-\gamma}) \to 0 \quad \text{as} \quad i\to\infty; \tag{20}$$

and

$$E\delta_i = O(i^{3\nu})O(i^{-\gamma}) \to 0 \quad \text{as} \quad i\to\infty. \tag{21}$$

Hence, from (16) – (21) we get,

$$E\left|\int_{\Re^k} V^*(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s - \int_{\Re^k} V^{\rho_i}_{n_i-1}(F(x,\hat{f}_i,s))\rho_i(s)\,\mathrm{d}s\right| \to 0 \quad \text{as} \quad i\to\infty, \tag{22}$$

which implies that

$$\liminf_{i\to\infty} E\int_{\Re^k} V^{\rho_i}_{n_i-1}(F(x,\hat{f}_i,s))\rho_i(s)\,\mathrm{d}s \geq \int_{\Re^k} V^*(F(x,f_\infty,s))\rho(s)\,\mathrm{d}s. \tag{23}$$

Indeed, for each $i\in\mathbb{N}$,

$$\int_{\Re^k} V^{\rho_i}_{n_i-1}(F(x,\hat{f}_i,s))\rho_i(s)\,\mathrm{d}s = \left[\int_{\Re^k} V^{\rho_i}_{n_i-1}(F(x,\hat{f}_i,s))\rho_i(s)\,\mathrm{d}s \right.$$
$$\left. - \int_{\Re^k} V^*(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s\right] + \int_{\Re^k} V^*(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s.$$

Now, taking expectation and lim inf as $i\to\infty$ on both sides of this equality, from (22) we get

$$\liminf_{i\to\infty} E\int_{\Re^k} V^{\rho_i}_{n_i-1}(F(x,\hat{f}_i,s))\rho_i(s)\,\mathrm{d}s \geq \liminf_{i\to\infty} E\int_{\Re^k} V^*(F(x,\hat{f}_i,s))\rho(s)\,\mathrm{d}s.$$

Thus, (23) follows from the lower semicontinuity of V^* (see Remark 4.5 (b)), the continuity of F in $a\in A(x)$, and Fatou's Lemma. Hence, taking expectation and lim inf as $i\to\infty$ in (14), and using the fact $\sup_{a\in A(x)} |c(x,a)-c_{n_i}(x,a)| \to 0$ (see Assumption 3.1), we obtain

$$c(x,f_\infty) + \alpha\int_{\Re^k} V^*(F(x,f_\infty,s))\rho(s)\,\mathrm{d}s \leq V^*(x). \tag{24}$$

Finally, as x was arbitrary, by (4), the equality holds in (24) for every $x\in X$, and therefore (see Proposition 3.2 (b)) $\{f_\infty\}$ is optimal for the model $\mathcal{M}$.

6. EXAMPLE

We consider a storage system of the form

$$x_{t+1} = (x_t + a_t - \xi_t)^+, \quad t = 0, 1, \ldots, \tag{25}$$

x_0 given, with state space $X = [0, \infty)$ and action set $A(x) = A = [0, \theta]$ for all $x \in X$ and some $\theta > 0$. In addition the random variables $\xi_0, \xi_1, \ldots$, are i.i.d. with a continuous and bounded density, satisfying

$$E[\xi_0] > \theta. \tag{26}$$

In particular, relation (25) describes an inventory-production system where x_t represents the stock level at the beginning of period t, the control a_t is the quantity ordered or produced at the beginning of period t, and the random variable ξ_t is the demand during that period.

Let Ψ be the moment generating function of the random variable $\theta - \xi_0$, that is, $\Psi(t) = E[\exp t(\theta - \xi_0)]$. Then, (26) implies $\Psi'(0) < 0$, and since $\Psi(0) = 1$, there exists $\lambda > 0$ such that

$$\beta_0 := \Psi(\lambda) < 1. \tag{27}$$

Now we fix a discount factor $\alpha = 1/2$, and let $c(x, a)$ be a nonnegative and continuous one-stage cost function such that

$$\sup_{a \in A} c(x, a) \leq \bar{c} e^{\lambda x},$$

for all $x \in X$ and some $\bar{c} > 0$. Defining $W(x) := \bar{c} e^{\lambda x}$, we have for all $x \in X$, $a \in A$,

$$\begin{aligned} \int_0^\infty \bar{c} e^{\lambda (x+a-s)^+} \rho(s)\, \mathrm{d}s &\leq \bar{c} + \int_0^\infty \bar{c} e^{\lambda (x+a-s)} \rho(s)\, \mathrm{d}s \\ &\leq \bar{c} + \bar{c} e^{\lambda x} \int_0^\infty e^{\lambda(\theta - s)} \rho(s)\, \mathrm{d}s \\ &= \bar{c} + \beta_0 \bar{c} e^{\lambda x} \leq \beta \bar{c} e^{\lambda x}, \end{aligned}$$

where $\beta := (1 + \beta_0)$. Observe that from (27) $\beta < 2$, and therefore Assumption 3.1 (a) is satisfied.

To verify Assumption 3.1 (c), let v be a bounded measurable function on X, and for every $a \in A(x)$, let ρ_a be the density of $a - \xi_0$. Observe that

$$\rho_a(y) = \rho(a - y), \quad -\infty < y \leq a.$$

In addition, for each $y \in \Re$, $a \to \rho_a(y)$ is continuous on A. Then,

$$\begin{aligned} \int_0^\infty v(x + y)^+ \rho_a(y)\, \mathrm{d}y &= v(0) \int_{-\infty}^{-x} \rho_a(y)\, \mathrm{d}y + \int_{-x}^\infty v(x + y) \rho_a(y)\, \mathrm{d}y \\ &= v(0) \int_{-\infty}^{-x} \rho_a(y)\, \mathrm{d}y + \int_0^\infty v(y) \rho_a(y - x)\, \mathrm{d}y. \end{aligned}$$

Thus, by Scheffé's Theorem,

$$a \to \int_0^\infty v[(x + a - s)^+]\rho(s)\,\mathrm{d}s$$

defines a continuous function on A. Finally, replacing v by the function W, and using similar arguments, we obtain that Assumption 3.1 (c) and (d) hold.

(Received December 17, 2003.)

REFERENCES

[1] R. Cavazos-Cadena: Nonparametric adaptive control of discounted stochastic systems with compact state space. J. Optim. Theory Appl. *65* (1990), 191–207.
[2] L. Devroye and L. Gyorfi: Nonparametric Density Estimation the L_1 View. Wiley, New York 1985.
[3] E. B. Dynkin and A. A. Yushkevich: Controlled Markov Processes. Springer–Verlag, New York 1979.
[4] E. I. Gordienko: Adaptive strategies for certain classes of controlled Markov processes. Theory Probab. Appl. *29* (1985), 504–518.
[5] E. I. Gordienko and J. A. Minjárez-Sosa: Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. Kybernetika *34* (1998), 217–234.
[6] R. Hasminskii and I. Ibragimov: On density estimation in the view of Kolmogorov's ideas in approximation theory. Ann. Statist. *18* (1990), 999–1010.
[7] O. Hernández-Lerma: Adaptive Markov Control Processes. Springer–Verlag, New York 1989.
[8] O. Hernández-Lerma and R. Cavazos-Cadena: Density estimation and adaptive control of Markov processes: average and discounted criteria. Acta Appl. Math. *20* (1990), 285–307.
[9] O. Hernández-Lerma and J. B. Lasserre: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer–Verlag, New York 1996.
[10] O. Hernández-Lerma and J. B. Lasserre: Further Topics on Discrete-Time Markov Control Processes. Springer–Verlag, New York 1999.
[11] O. Hernández-Lerma and S. I. Marcus: Adaptive policies for discrete-time stochastic control systems with unknown disturbance distribution. Systems Control Lett. *9* (1987), 307–315.
[12] N. Hilgert and J. A. Minjárez-Sosa: Adaptive policies for time-varying stochastic systems under discounted criterion. Math. Methods Oper. Res. *54* (2001), 491–505.
[13] M. Schäl: Conditions for optimality and for the limit of n-stage optimal policies to be optimal. Z. Wahrsch. Verw. Gerb. *32* (1975), 179–196.

J. Adolfo Minjárez-Sosa, Departamento de Matemáticas, Universidad de Sonora, Rosales s/n, Col. Centro, 83000, Hermosillo, Sonora. México.
e-mail: aminjare@gauss.mat.uson.mx