Morfismos, Vol. 12, No. 2, 2008, pp. 33–52

Asymptotic normality of average cost Markov control processes *

Armando F. Mendoza-Pérez

Abstract

This paper studies asymptotic normality of Markov control processes (MCPs) in Borel spaces with unbounded cost. Under suitable hypotheses we show that within the class of canonical policies there exists one where the cost is asymptotically normal.

2000 Mathematics Subject Classification: 93E20, 90C40.
Keywords and phrases: (discrete-time) Markov control processes, average cost criteria, expected average cost, average variance, asymptotic normality.

1 Introduction.

We study the asymptotic normality of discrete-time MCPs in Borel spaces with possibly unbounded cost. Under suitable hypotheses we show that within the class of so-called canonical policies, those that minimize the limiting *average variance* have an asymptotic normality behavior, that is, certain distribution of the cost is asymptotically normal. Asymptotic normality is very useful in adaptive control problems.

The only works for the variance minimization problem in MCPs are those by Mandl [7, 9, 10], Hernández-Lerma et al. [5], Prieto-Rumeau and Hernández-Lerma [11] and Zhu and Guo [15]. For the asymptotic behavior of the MCPs, there are a lot fewer works. For instance, we should mention the paper by Mandl [8] for finite state MCPs.

*This paper is part of the author's Doctoral Thesis written at the Departamento de Matemáticas, CINVESTAV-IPN.

To obtain our results we combine two approaches. The first one, to obtain canonical policies with minimum average variance, we use the W-uniform ergodicity assumptions in [5]. The second one follows Mandl's approach [8] to extend asymptotic normality for MCPs in Borel spaces.

The remainder of the paper is organized as follow. Section 2 contains a brief description of the Markov control model of interest. In Section 3 we introduce our hypotheses and state our main result, Theorem 3.7, which is proved in Section 4. Finally, a LQ system in Section 5 illustrates our results.

2 The control model.

Let $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C)$ be a discrete time Markov control model with state space $\mathbf{X}$ and control (or action) set $\mathbf{A}$, both assumed to be Borel spaces with σ-algebras $\mathcal{B}(\mathbf{X})$ and $\mathcal{B}(\mathbf{A})$, respectively. For each $x \in \mathbf{X}$ there is a nonempty Borel set $A(x)$ in $\mathcal{B}(\mathbf{A})$ which represents the set of feasible actions in the state x. The set

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

is assumed to be a Borel subset of $\mathbf{K} \times \mathbf{A}$. The transition law Q is a stochastic kernel on $\mathbf{X}$ given $\mathbf{K}$ and the one-stage cost C is a real-valued measurable function on $\mathbf{K}$.

The class of measurable functions $f : \mathbf{X} \to \mathbf{A}$ such that $f(x)$ is in $A(x)$ for every $x \in \mathbf{X}$ is denoted by $\mathbf{F}$ and we suppose that is nonempty.

Control policies. For every $n = 0, 1, \ldots$, let $\mathbf{H}_n$ be the family of admissible histories up to time n; that is, $\mathbf{H}_0 := \mathbf{X}$, and $\mathbf{H}_n := \mathbf{K}^n \times \mathbf{X}$ if $n \geq 1$. A *control policy* is a sequence $\pi = \{\pi_n\}$ of stochastic kernels π_n on $\mathbf{A}$ given $\mathbf{H}_n$ such that $\pi_n(A(x_n)|h_n) = 1$ for every n-history $h_n = (x_0, a_0, \cdots, x_{n-1}, a_{n-1}, x_n)$ in $\mathbf{H}_n$. The class of all policies is denoted by Π.

A policy $\pi = \{\pi_n\}$ is said to be a (deterministic) *stationary policy* if there exists $f \in \mathbf{F}$ such that $\pi_n(\cdot|h_n)$ is the Dirac measure at $f(x_n) \in A(x_n)$ for all $h_n \in \mathbf{H}_n$ and $n = 0, 1, \ldots$. Following a standard convention, we identify $\mathbf{F}$ with the class of stationary policies.

For notational ease we write

$$(1) \quad C_f(x) := C(x, f(x)) \quad \text{and} \quad Q_f(\cdot|x) := Q(\cdot|x, f(x)) \quad \forall x \in \mathbf{X}$$

for every stationary policy f in $\mathbf{F}$.

Let $(\Omega, \mathcal{F})$ be the (canonical) measurable space consisting of the sample space $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$ and its product σ-algebra $\mathcal{F}$. Then, for each policy π and "initial state" $x \in \mathbf{X}$, a stochastic process $\{(x_n, a_n)\}$ and a probability measure P_x^π are defined on $(\Omega, \mathcal{F})$ in a canonical way, where x_n and a_n represent the state and control at time n, $n = 0, 1, \ldots$. The expectation operator with respect to P_x^π is denoted by E_x^π.

Average cost criteria. For each $n = 1, 2, \ldots$, let

$$J_n(\pi, x) := E_x^\pi \sum_{t=0}^{n-1} C(x_t, a_t)$$

be the *n-stage expected cost* when using the policy π, given the initial state $x \in \mathbf{X}$. The long-run *expected average cost* (EAC) is then defined as

$$J(\pi, x) := \limsup_{n\to\infty} \frac{1}{n} J_n(\pi, x). \tag{2}$$

Definition 2.1 *(a) A policy π^* is said to be EAC-optimal if*

$$J(\pi^*, x) = \inf_{\pi\in\Pi} J(\pi, x) =: J^*(x) \quad \forall x \in \mathbf{X}. \tag{3}$$

(b) A stationary policy $f_* \in \mathbf{F}$ is called canonical if there exists a constant ρ_* and a measurable function $h_1 : \mathbf{X} \to \mathbf{R}$ such that

$$\rho_* + h_1(x) = \min_{a\in A(x)} \left[C(x, a) + \int_{\mathbf{X}} h_1(y) Q(dy|x, a) \right] \quad \forall x \in \mathbf{X}, \tag{4}$$

and $f_*(x) \in A(x)$ attain the minimum on the right-hand side of (4) for every $x \in \mathbf{X}$, i.e.,

$$\rho_* + h_1(x) = C_{f_*}(x) + \int_{\mathbf{X}} h_1(y) Q_{f_*}(dy|x) \quad \forall x \in \mathbf{X}. \tag{5}$$

If (4) and (5) are satisfied, then (ρ_*, h_1, f_*) is said to be a canonical triplet (see [1, 2, 14]).

Remark 2.2 *(See [2, Section 5.2].) If (ρ_*, h_1, f_*) is a canonical triplet and in addition h_1 satisfies that*

$$\lim_{n\to\infty} \frac{1}{n} E_x^\pi h_1(x_n) = 0 \quad \forall \pi \in \Pi, x \in \mathbf{X}, \tag{6}$$

then f_* is EAC-optimal and ρ_* is the optimal expected average cost, that is,

$$J(f_*, x) = J^*(x) = \rho_* \quad \forall x \in \mathbf{X}. \tag{7}$$

Hence we have

$$\mathbf{F}_{cp} \subset \mathbf{F}_{eac}, \tag{8}$$

where $\mathbf{F}_{cp}$ *is the class of canonical policies and* $\mathbf{F}_{eac} \subset \mathbf{F}$ *is the class of stationary EAC-optimal policies.*

For each $n = 1, 2, \ldots$, let

$$S_n(f,x) := \sum_{t=0}^{n-1} C(x_t, a_t) \tag{9}$$

be the *n-stage pathwise (or sample-path) cost* when using the policy $f \in \mathbf{F}$, given the initial state $x \in \mathbf{X}$.

Definition 2.3 *(a) For each* $f \in \mathbf{F}$ *and* $x \in \mathbf{X}$*, define the limiting average variance*

$$V(f,x) := \limsup_{n\to\infty} \frac{1}{n} E_x^f \left[S_n(f,x) - J_n(f,x) \right]^2. \tag{10}$$

(b) A stationary policy $\hat{f}$ *is called variance-minimal if*

$$V(\hat{f},x) = \inf_{f\in\mathbf{F}_{eac}} V(f,x) \quad \forall x \in \mathbf{X}. \tag{11}$$

3 Assumptions and main result.

In this section we introduce conditions to study asymptotic normality.

We shall first introduce two sets of hypotheses. The first one, Assumption 3.1, consists of standard continuity-compactness conditions (see, for instance, [1, 3, 5, 12]) together with a growth condition on the one-step cost C.

Assumption 3.1 *For every state* $x \in \mathbf{X}$*:*

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

(b) *C(x,a) is lower semicontinuous in* $a \in A(x)$*;*

(c) *the function* $a \mapsto \int_{\mathbf{X}} u(y)Q(dy|x,a)$ *is continuous on* $A(x)$ *for every bounded measurable function* u *on* $\mathbf{X}$*;*

(d) *there exists a measurable function* $W \geq 1$*, a bounded measurable function* $b \geq 0$*, and nonnegative constants* r_1 *and* β *with* $\beta < 1$*, such that* (d1) $|C(x,a)| \leq r_1 W(x) \quad \forall (x,a) \in \mathbf{K}$ *and*

(d2) $\int_{\mathbf{X}} W(y) Q(dy|x,a)$ *is continuous in* $a \in A(x)$*; and*

(d3) $\int_{\mathbf{X}} W(y) Q(dy|x,a) \leq \beta W(x) + b(x)$ *for every* $x \in \mathbf{X}$.

To state our second set of hypotheses, let us first introduce the following notation: $B_W(\mathbf{X})$ denotes the normed linear space of measurable functions u on $\mathbf{X}$ with finite W-norm $\|u\|_W$, which is defined as

$$\|u\|_W := \sup_{x \in \mathbf{X}} |u(x)|/W(x). \tag{12}$$

In this case we say that u is W-bounded.

Let $\mu(\cdot)$ be a measure on $\mathbf{X}$. We write

$$\mu(u) := \int_{\mathbf{X}} u(y) \mu(dy) \tag{13}$$

whenever the integral is well-defined.

Assumption 3.2 *For each stationary policy* $f \in \mathbf{F}$*:*

(a) *(W-geometric ergodicity) There exists a probability measure* μ_f *on* $\mathbf{X}$ *such that*

$$\left| \int_{\mathbf{X}} u(y) Q_f^t(dy|x) - \mu_f(u) \right| \leq \|u\|_W R \rho^t W(x), \tag{14}$$

for every $t = 0, 1, \ldots,$ u *in* $B_W(\mathbf{X})$ *and* $x \in \mathbf{X}$*, where* $R > 0$ *and* $0 < \rho < 1$ *are constants independent of* f.

(b) *(Irreducibility) There exists a* σ*-finite measure* λ *on* $\mathcal{B}(\mathbf{X})$ *with respect to which* Q_f *is* λ*-irreducible.*

Remark 3.3 *(See [4, Theorem 3.5],[13, Theorem 4.5.3],[3, Theorem 10.3.6].) Under Assumptions 3.1 and 3.2, there exists a canonical triplet* (ρ_*, h_1, f_*)*; see Definition 2.1.*

To obtain asymptotic normality we need to strengthen the growth condition on the cost function C in Assumption 3.1(d1).

Assumption 3.4 *There exists a positive constant* r_2 *such that*

$$C^4(x,a) \leq r_2 W(x) \quad \forall (x,a) \in \mathbf{K}. \tag{15}$$

Remark 3.5 (a) *Because* $W \geq 1$*, Assumption 3.4 implies Assumption 3.1(d1). Moreover, we have that* $C^2(x,a) \leq {r_2}^{1/2}W(x)$ *for every* (x,a) *in* $\mathbf{K}$ *(Assumption 3.6 in [5]), condition which is necessary to obtain optimal policies with minimal average variance.*

(b) *Under Assumptions 3.1, 3.2 and 3.4, the function* h_1 *satisfying (4) and (5) above is such that* h_1^2 *and* h_1^4 *belong to* $B_W(\mathbf{X})$*. (See Lemma 4.3 below.)*

By the Remark 3.5(b), the function $\Lambda(\cdot,\cdot)$ on $\mathbf{K}$ defined as

$$\Lambda(x,a) := \int_{\mathbf{X}} {h_1}^2(y)Q(dy|x,a) - \left[\int_{\mathbf{X}} h_1(y)Q(dy|x,a)\right]^2 \tag{16}$$

is finite-valued. This function is used to state the following variance-minimization result.

Proposition 3.6 *(See [5, Theorem 3.8] or [3, Theorem 11.3.8].) Under Assumptions 3.1, 3.2 and 3.4, there exists a constant* $\sigma_*^2 \geq 0$*, a deterministic canonical policy* $f_* \in \mathbf{F}_{cp}$*, and a function* h_2 *in* $B_W(\mathbf{X})$ *such that, for each* $x \in \mathbf{X}$*,*

$$\sigma_*^2 + h_2(x) = \Lambda_{f_*}(x) + \int_{\mathbf{X}} h_2(y)Q_{f_*}(dy|x) \tag{17}$$

Furthermore, f_* *satisfies (11) and* $V(f_*,\cdot) = \sigma_*^2$*; in fact*

$$V(f_*,x) = \mu_{f_*}(\Lambda_{f_*}) = \sigma_*^2 \quad \forall x \in \mathbf{X} \tag{18}$$

and

$$\sigma_*^2 \leq V(f,x) \quad \forall f \in \mathbf{F}_{eac}, x \in \mathbf{X}. \tag{19}$$

Hence, (19) states that σ_*^2 is the minimal average variance. We can now state our main result, which is proved in Section 4.

Theorem 3.7 *Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Let* $f_* \in \mathbf{F}_{cp}$ *be a canonical policy satisfying Proposition 3.6, and* ρ_* *the optimal average cost as in (7). Then for every initial state* $x \in \mathbf{X}$*,*

$$\frac{S_n(f_*,x) - n\rho_*}{\sqrt{n}} \tag{20}$$

has asymptotically a normal distribution $N(0,\sigma_*^2)$ *as* $n \to \infty$*, with* $S_n(f_*,x)$ *as in (9).*

4 Proof of Theorem 3.7.

In the remainder of this paper we suppose that Assumptions 3.1, 3.2 and 3.4 hold.

To prove Theorem 3.7 we need some preliminary results, which are stated as Lemmas 4.1, 4.2, 4.3.

The following lemma summarizes some well-known results, which are stated here for ease of reference.

Lemma 4.1 *Let $f \in \mathbf{F}$ be a deterministic stationary policy and $\{x_t\}$ the Markov chain induced by f. Then*

(a) *[3, Lemma 10.4.1] For each $x \in \mathbf{X}$ and $t = 1, 2, \ldots$*

$$E_x^f W(x_t) \leq [1 + b/(1-\beta)]W(x), \tag{21}$$

with $b := \sup_{x\in\mathbf{X}} |b(x)|$. Moreover, for every function u in $B_W(\mathbf{X})$ the following limits hold:

$$\lim_{n\to\infty} \frac{1}{n^p} E_x^f u(x_n) = 0 \tag{22}$$

with $p > 0$.

(b) *[3, Proposition 10.2.3] $|J_n(f,x) - nJ_f| \leq r_1 RW(x)/(1-\rho) \quad \forall x \in \mathbf{X}, n = 1, 2, \ldots$, where $J_f := \mu_f(C_f)$. Hence:*

(c) *$J(f,x) = \lim_{n\to\infty} J_n(f,x)/n = J_f \quad \forall x \in \mathbf{X}$.*

(d) *[3, Proposition 10.2.3] The function*

$$\begin{aligned} h_f(x) &:= \lim_{n\to\infty} [J_n(f,x) - nJ_f] \\ &= \sum_{t=0}^{\infty} E_x^f [C_f(x_t) - J_f] \end{aligned} \tag{23}$$

belongs to $B_W(\mathbf{X})$ which is called the "bias of f". Moreover, by (b), we have

$$\|h_f\|_W \leq r_1 R/(1-\rho). \tag{24}$$

(e) *[3, Theorem 10.3.6] The pair (J_f, h_f) is the unique solution of the Poisson equation*

$$J_f + h_f(x) = C_f(x) + \int_{\mathbf{X}} h_f(y) Q_f(dy|x), \forall x \in \mathbf{X}, \tag{25}$$

that satisfies the condition $\mu_f(h_f) = 0$.

(f) *[3, Theorem 10.3.7] If f is a canonical policy in $\mathbf{F}_{cp}$, the corresponding solution $(J_f, h_f) = (\rho_*, h_f)$ to the Poisson equation (25) is such that h_f coincides with the function h_1, with h_1 as in (4) and (5), that is,*

$$h_f(\cdot) = h_1(\cdot) + k_f$$

for some constant k_f.

The following lemma states a stronger version of (14) and Lemma 4.1(e).

Lemma 4.2 *Let $w(x) := W(x)^{1/m}$ with $m = 2$ or $m = 4$. For each stationary policy $f \in \mathbf{F}$:*

(a) *The Markov chain $\{x_n\}$ induced by f is w-geometrically ergodic, that is,*

$$\left| \int_{\mathbf{X}} u(y) Q_f^t(dy|x) - \mu_f(u) \right| \leq \|u\|_w R_0 \rho_0^t w(x) \tag{26}$$

for all $x \in \mathbf{X}$ and $t = 0, 1, \ldots$, where $\rho_0 = \rho^{1/m} < 1$ and $R_0 := R^{1/m}$;

(b) *The unique solution (J_f, h_f) of the Poisson equation (25) is such that h_f is w-bounded.*

Proof. **(a)** This part follows from [3, Lemma 11.3.9].

(b) Case $m = 4$: Note that (15) and part (a) of this lemma yield the $W^{1/4}$-analogue of Lemma 4.1(d). Hence h_f is $W^{1/4}$-bounded.

Case $m = 2$: Assumption 3.4 and the fact that $W \geq 1$ imply that

$$|C(x,a)| \leq r_2^{1/4} W(x)^{1/4} \leq r_2^{1/4} W(x)^{1/2} \quad \forall (x,a) \in \mathbf{K}. \tag{27}$$

Part (a) (with $m = 2$) and (27) yield the $W^{1/2}$-analogue of Lemma 4.1(d), that is, h_f is $W^{1/2}$-bounded. □

Lemma 4.3 (a) *The function $h_1(\cdot)$ satisfying (4) and (5) is $W^{1/4}$-bounded.*

(b) *The function $h_2(\cdot)$ satisfying (17) is $W^{1/2}$-bounded.*

Proof. **(a)** By Lemma 4.1(f), h_1 coincides with h_f except for an additive constant, with f a canonical policy. From Lemma 4.2(b), h_f is $W^{1/4}$-bounded, therefore h_1 is also $W^{1/4}$-bounded.

(b) From the proof of Proposition 3.6 (see for instance, [5, Theorem 3.8] or [3, Theorem 11.3.8]) we consider the new Markov control model

$$(\mathbf{X}, \mathbf{A}, \{A^*(x) : x \in \mathbf{X}\}, Q, \Lambda), \tag{28}$$

with $A^*(x)$ an appropriate compact subset of $A(x)$ for every x, and $\Lambda(x, a)$ as in (16). From part (a) of this lemma, h_1 is $W^{1/4}$-bounded. Hence we have that Λ satisfies the following growth condition

$$\Lambda^2(x, a) \leq r_3 W(x) \quad \forall (x, a) \in \mathbf{K}, \tag{29}$$

where r_3 is a positive constant. Observe that (29) yields the $W^{1/2}$-analogue of Assumption 3.1(d1); hence, by Lemma 4.2(a), the control model (28) is $W^{1/2}$-geometrically ergodic. Then from Lemma 4.1 applied to the control model (28) with $W^{1/2}$ instead of W, and h_2 instead of h_1, it follows that h_2 is $W^{1/2}$-bounded. □

We are finally ready for the proof of Theorem 3.7.

Proof of Theorem 3.7. Let (ρ_*, h_1, f_*) be a canonical triplet as in Definition 2.1. Moreover, let (σ_*^2, h_2, f_*) be as in Proposition 3.6.

We define

$$\tau_1(x, a) := \int_{\mathbf{X}} h_1(y) Q(dy|x, a) - h_1(x) + C(x, a) - \rho^*$$

and

$$\tau_2(x, a) := \int_{\mathbf{X}} h_2(y) Q(dy|x, a) - h_2(x) + \Lambda(x, a) - \sigma_*^2$$

for all $(x, a) \in \mathbf{K}$. For $l = 1, 2$, and $x \in \mathbf{X}$, let

$$\psi_l(x, a) := \int_{\mathbf{X}} h_l(y) Q(dy|x, a) - h_l(x),$$

and consider the characteristic functions

$$\chi_n(u) := \exp\{iu(S_n(f_*, x) - n\rho_*)\} \quad \text{for } n = 1, 2, \cdots; u \in \mathbf{R},$$

with $\chi_0(u) := 1$. Let

$$e_1(z) := \exp\{iz\} - iz - 1, \tag{30}$$

$$e_2(z) := \exp\{iz\} + \frac{z^2}{2} - iz - 1. \tag{31}$$

Observe that

$$\tau_1(x,a) = \psi_1(x,a) + C(x,a) - \rho_*, \tag{32}$$

and

$$\tau_2(x,a) = \psi_2(x,a) + \Lambda(x,a) - \sigma_*^2 \tag{33}$$

for all $(x,a) \in \mathbf{K}$.

To prove the theorem we have to verify that

$$\lim_{n\to\infty} E_x^{f_*} \chi_n\Big(\frac{u}{\sqrt{n}}\Big) = \exp\{-\frac{1}{2}\sigma_*^2 u^2\}. \tag{34}$$

To this end, first notice that $\psi_l(x_m, a_m)$ for $l = 1, 2$, is the conditional expectation of $h_l(x_{m+1}) - h_l(x_m)$ given x_m, a_m, that is,

$$\psi_l(x_m, a_m) = E_x^{f_*}[h_l(x_{m+1}) - h_l(x_m)|x_m, a_m].$$

This yields for $l = 1, 2$, with $\chi_m := \chi_m(u)$ and $\psi_l := \psi_l(x_m, a_m)$, the equations

$$0 = iuE_x^{f_*}\left[\sum_{m=0}^{n-1} \chi_m \psi_1 - \sum_{m=0}^{n-1} \chi_m \Big(h_1(x_{m+1}) - h_1(x_m)\Big)\right] \tag{35}$$

and

$$0 = \frac{u^2}{2} E_x^{f_*}\left[\sum_{m=0}^{n-1} \chi_m \Big(h_2(x_{m+1}) - h_2(x_m)\Big) - \sum_{m=0}^{n-1} \chi_m \psi_2\right]. \tag{36}$$

To simplify the notation, let $C := C(x_m, a_m)$, $e_1 := e_1\Big(u(C - \rho_*)\Big)$ and $e_2 := e_2\Big(u(C - \rho_*)\Big)$. Moreover, notice that

$$\chi_{m+1} - \chi_m = \Big[\exp\{iu(C - \rho_*)\} - 1\Big]\chi_m. \tag{37}$$

From (30), (31) and (37) we have

$$\begin{aligned} E_x^{f_*}\chi_n - 1 &= E_x^{f_*} \sum_{m=0}^{n-1} (\chi_{m+1} - \chi_m) \\ &= E_x^{f_*} \sum_{m=0}^{n-1} \Big[iu(C - \rho_*) - \frac{1}{2}u^2(C - \rho_*)^2 + e_2\Big]\chi_m, \end{aligned} \tag{38}$$

and

$$-iuE_x^{f_*} \sum_{m=0}^{n-1} \chi_m \Big(h_1(x_{m+1}) - h_1(x_m)\Big) =$$

$$iuE_x^{f_*}\Big[h_1(x_0)-\chi_n h_1(x_n)+\sum_{m=0}^{n-1}h_1(x_{m+1})(\chi_{m+1}-\chi_m)\Big]=$$

$$iuE_x^{f_*}\Big[h_1(x_0)-\chi_n h_1(x_n)+$$

$$\sum_{m=0}^{n-1}h_1(x_{m+1})\Big(iu(C-\rho_*)+e_1\Big)\chi_m\Big]. \tag{39}$$

Similarly,

$$\frac{u^2}{2}E_x^{f_*}\sum_{m=0}^{n-1}\chi_m\Big(h_2(x_{m+1})-h_2(x_m)\Big)=$$

$$-\frac{u^2}{2}E_x^{f_*}\Big[h_2(x_0)-\chi_n h_2(x_n)+\sum_{m=0}^{n-1}h_2(x_{m+1})(\chi_{m+1}-\chi_m)\Big]=$$

$$-\frac{u^2}{2}E_x^{f_*}\Big[h_2(x_0)-\chi_n h_2(x_n)+$$

$$\sum_{m=0}^{n-1}h_2(x_{m+1})\Big(\exp\{iu(C-\rho_*)\}-1\Big)\chi_m\Big]. \tag{40}$$

Adding (35)-(40) and using (32)

$$E_x^{f_*}\chi_n-1$$

$$=iuE_x^{f_*}\Big[h_1(x_0)-\chi_n h_1(x_n)+\sum_{m=0}^{n-1}\chi_m\tau_1(x_m,a_m)+\sum_{m=0}^{n-1}e_1h_1(x_{m+1})\chi_m\Big]$$

$$-\frac{u^2}{2}E_x^{f_*}\sum_{m=0}^{n-1}\chi_m\Big\{\psi_2+2h_1(x_{m+1})(C-\rho_*)+(C-\rho_*)^2\Big\}$$

$$-\frac{u^2}{2}E_x^{f_*}\Big[h_2(x_0)-\chi_n h_2(x_n)+\sum_{m=0}^{n-1}h_2(x_{m+1})\Big(\exp\{iu(C-\rho_*)\}-1\Big)\chi_m\Big]$$

$$+E_x^{f_*}\sum_{m=0}^{n-1}e_2\chi_m.$$

Hence

$$E_x^{f_*}\chi_n-1=\kappa''(n,u)-$$

$$\frac{u^2}{2}E_x^{f_*}\sum_{m=0}^{n-1}\chi_m\Big\{\psi_2+2h_1(x_{m+1})(C-\rho_*)+(C-\rho_*)^2\Big\} \tag{41}$$

with

$$\kappa''(n,u) =$$

$$iuE_x^{f_*}\Big[h_1(x_0) - \chi_n h_1(x_n) + \sum_{m=0}^{n-1} \chi_m \tau_1(x_m, a_m) + \sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big]$$

$$-\frac{u^2}{2}E_x^{f_*}\Big[h_2(x_0) - \chi_n h_2(x_n) + \sum_{m=0}^{n-1} h_2(x_{m+1})\Big(\exp\{iu(C-\rho_*)\}-1\Big)\chi_m\Big]$$

$$+E_x^{f_*}\sum_{m=0}^{n-1} e_2\chi_m. \tag{42}$$

Observing that

$$\Lambda(x_m, a_m) = E_x^{f_*}[h_1^2(x_{m+1})|x_m, a_m] - \Big(E_x^{f_*}[h_1(x_{m+1})|x_m, a_m]\Big)^2$$

and in view of (33), we can express (41) as

$$E_x^{f_*}\chi_n - 1$$

$$= \kappa''(n,u) - \frac{u^2}{2}E_x^{f_*}\sum_{m=0}^{n-1}\chi_m\Big\{\sigma_*^2 + \tau_2(x_m, a_m) - h_1^2(x_{m+1})$$

$$+\Big(E_x^{f_*}[h_1(x_{m+1})|x_m, a_m] + C(x_m, a_m) - \rho_*\Big)^2\Big\}$$

$$= \kappa''(n,u) - \frac{u^2}{2}E_x^{f_*}\sum_{m=0}^{n-1}\chi_m\Big\{\sigma_*^2 + \tau_2(x_m, a_m) - h_1^2(x_{m+1})$$

$$+\Big(\int_{\mathbf{X}} h_1(y)Q(dy|x_m, a_m) + C(x_m, a_m) - \rho_*\Big)^2\Big\}.$$

Since f_* is a canonical policy, it satisfies

$$h_1(x_m) = \int_{\mathbf{X}} h_1(y)Q(dy|x_m, a_m) + C(x_m, a_m) - \rho_*.$$

Then, from (37), we have

$$E_x^{f_*}\chi_n - 1$$

$$= \kappa''(n,u) - \frac{u^2}{2}E_x^{f_*}\sum_{m=0}^{n-1}\chi_m\Big\{\sigma_*^2 + \tau_2(x_m, a_m) - h_1^2(x_{m+1}) + h_1^2(x_m)\Big\}$$

$$= \kappa''(n,u) - \frac{u^2\sigma_*^2}{2}\sum_{m=0}^{n-1} E_x^{f_*}\chi_m - \frac{u^2}{2}E_x^{f_*}\Big[h_1^2(x_0) - \chi_n h_1^2(x_n)$$

$$+ \sum_{m=0}^{n-1}\chi_m\tau_2(x_m,a_m) + \sum_{m=0}^{n-1} h_1^2(x_{m+1})(\chi_{m+1}-\chi_m)\Big].$$

$$= \kappa''(n,u) - \frac{u^2\sigma_*^2}{2}\sum_{m=0}^{n-1} E_x^{f_*}\chi_m - \frac{u^2}{2}E_x^{f_*}\Big[h_1^2(x_0) - \chi_n h_1^2(x_n)$$

$$+ \sum_{m=0}^{n-1}\chi_m\tau_2(x_m,a_m) + \sum_{m=0}^{n-1} h_1^2(x_{m+1})\Big(\exp\{iu(C-\rho_*)\}-1\Big)\chi_m\Big].$$

Hence

$$E_x^{f_*}\chi_n = 1 - \frac{u^2\sigma_*^2}{2}\sum_{m=0}^{n-1} E_x^{f_*}\chi_m + \kappa'(n,u) \tag{43}$$

with

$$\kappa'(n,u) = \kappa''(n,u) - \frac{u^2}{2}E_x^{f_*}\Big[h_1^2(x_0) - \chi_n h_1^2(x_n) + \sum_{m=0}^{n-1}\chi_m\tau_2(x_m,a_m)$$

$$+ \sum_{m=0}^{n-1} h_1^2(x_{m+1})\Big(\exp\{iu(C-\rho_*)\}-1\Big)\chi_m\Big]. \tag{44}$$

Let us rewrite (43) as

$$E_x^{f_*}\chi_n = 1 + \Big(\exp\{-\frac{u^2\sigma_*^2}{2}\}-1\Big)\sum_{m=0}^{n-1} E_x^{f_*}\chi_m + \kappa(n,u), \tag{45}$$

with

$$\kappa(n,u) := \kappa'(n,u) + \Big[1 - \frac{u^2\sigma_*^2}{2} - \exp\{-\frac{u^2\sigma_*^2}{2}\}\Big]\sum_{m=0}^{n-1} E_x^{f_*}\chi_m. \tag{46}$$

From (45), an induction argument gives

$$E_x^{f_*}\chi_n(u) = \exp\{-\frac{n\sigma_*^2u^2}{2}\} +$$

$$\Big[\exp\{-\frac{\sigma_*^2u^2}{2}\}-1\Big]\sum_{m=0}^{n-1}\exp\Big\{-\frac{\sigma_*^2u^2}{2}(n-1-m)\Big\}\kappa(m,u)$$

$$+\kappa(n,u). \tag{47}$$

Observe that the proof of the limit (34) and consequently of Theorem 3.7 follows from (47) if we show

$$\max_{1\le m\le n} |\kappa(m, \frac{u}{\sqrt{n}})| \to 0 \quad \text{as} \quad n \to \infty. \tag{48}$$

This relation is obtained by an inspection of the different terms of $\kappa(m, u/\sqrt{n})$. We will do this in the following six steps.

(i) Since f_* is a canonical policy satisfying (5), we have $\tau_1(x_m, a_m) = 0$ for $m = 0, 1, \cdots$ in (42). Similarly, by (17), $\tau_2(x_m, a_m) = 0$ in (44).

(ii) From (22) we have that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} E_x^{f_*} h(x_n) = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} E_x^{f_*} h(x_n) = 0$$

for every h in $B_W(\mathbf{X})$. This limit appears in (42) and (44) when we replace u by $u/\sqrt{n}$.

(iii) In this part we prove the limit (see (42))

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} E_x^{f_*} \sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m = 0.$$

From the fact $|e_1(z)| \le z^2/2$ for all z in $\mathbf{R}$, we obtain

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{n}} E_x^{f_*} \sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big| \\
&\quad\le \frac{1}{2\sqrt{n}} E_x^{f_*} \sum_{m=0}^{n-1} \frac{u^2}{n} |h_1(x_{m+1})|(C(x_m, a_m) - \rho_*)^2 \\
&\quad= \frac{u^2}{2n^{3/2}} E_x^{f_*} \sum_{m=0}^{n-1} \Big|\int_{\mathbf{X}} h_1(y) Q_{f_*}(dy|x_m)\Big|(C_{f_*}(x_m) - \rho_*)^2.
\end{aligned}$$

By Lemma 4.3(a), $h_1(\cdot)$ is $\sqrt[4]{W}$-bounded, in particular $h_1(\cdot)$ is $\sqrt{W}$-bounded. Hence the function $\int_{\mathbf{X}} h_1(y) Q_{f_*}(dy|\cdot)$ is $\sqrt{W}$-bounded. On the other hand, by Assumption 3.4 $(C_{f_*}(x) - \rho_*)^2$ is also $\sqrt{W}$-bounded. Therefore

$$\Big|\frac{1}{\sqrt{n}} E_x^{f_*} \sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big| \le \frac{\lambda u^2}{2n^{3/2}} E_x^{f_*} \sum_{m=0}^{n-1} W(x_m)$$

where λ is a constant depending on h_1 and C. By (21) we obtain

$$\Big|\frac{1}{\sqrt{n}} E_x^{f_*} \sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big| \le \frac{\lambda u^2}{2n^{3/2}} n[1 + b/(1-\beta)]W(x).$$

which converges to zero as $n \to \infty$.

(iv) We shall next prove

$$\lim_{n\to\infty} \frac{1}{n} E_x^{f_*} \sum_{m=0}^{n-1} e_2 \chi_m = 0.$$

This limit appears in (42) when we replace u by $u/\sqrt{n}$.

Observe that $|e_2(z)| \leq |z|^3/6$ for all z in $\mathbf{R}$. So, by Assumptions 3.1(d) and 3.4, together with (21),

$$\begin{aligned}
\Big|\frac{1}{n} E_x^{f_*} \sum_{m=0}^{n-1} e_2 \chi_m\Big| &\leq \frac{|u|^3}{6n^{5/2}} E_x^{f_*} \sum_{m=0}^{n-1} |C_{f_*}(x_m) - \rho_*|^3 \\
&\leq \frac{k^3|u|^3}{6n^{5/2}} E_x^{f_*} \sum_{m=0}^{n-1} W(x_m)^{3/4} \\
&\leq \frac{k^3|u|^3}{6n^{5/2}} E_x^{f_*} \sum_{m=0}^{n-1} W(x_m) \\
&\leq \frac{k^3|u|^3}{6n^{3/2}} [1 + b/(1-\beta)] W(x)
\end{aligned}$$

which converges to zero as $n \to \infty$, with k a constant.

(v) Let h be a $\sqrt{W}$-bounded function on $\mathbf{X}$. Then

$$\lim_{n\to\infty} \frac{1}{n} E_x^{f_*} \sum_{m=0}^{n-1} h(x_{m+1}) \Big(\exp\{i \frac{u}{\sqrt{n}} (C - \rho_*)\} - 1 \Big) \chi_m = 0.$$

This limit appears in (42) and (44) when u is replaced by $u/\sqrt{n}$.

It follows from the relation $e_1(z) = \exp\{iz\} - iz - 1$ that

$$\exp\{i \frac{u}{\sqrt{n}} (C - \rho_*)\} - 1 = i \frac{u}{\sqrt{n}} (C - \rho_*) + e_1 \Big(\frac{u}{\sqrt{n}} (C - \rho_*) \Big).$$

So

$$|\frac{1}{n} E_x^{f_*} \sum_{m=0}^{n-1} h(x_{m+1}) \Big(\exp\{i \frac{u}{\sqrt{n}} (C - \rho_*)\} - 1 \Big) \chi_m | \leq$$

$$\frac{|u|}{n^{3/2}} E_x^{f_*} \sum_{m=0}^{n-1} |h(x_{m+1})| |(C_{f_*}(x_m) - \rho_*)| + \frac{1}{n} E_x^{f_*} \sum_{m=0}^{n-1} |h(x_{m+1})| |e_1|.$$

This gives the desired conclusion by similar arguments to those in (iii).

(vi) The absolute value of the expression within brackets in (46) is majorized by $\sigma_*^4 u^4/8$, then the corresponding term in $\kappa(n, u/\sqrt{n})$ is majorized by $\sigma_*^4 u^4/8n^2$.

The statements (i)-(vi) imply (48) and consequently prove the theorem. □

Remark 4.4 *Taking* **A** *as a single-point set (singleton) we obtain the Central Limit Theorem for (noncontrolled) Markov chains.*

5 An example: a LQ system

Consider the linear system

$$x_{t+1} = k_1 x_t + k_2 a_t + z_t, \quad t = 0, 1, \cdots, \tag{49}$$

with state space $\mathbf{X} := \mathbf{R}$ and positive coefficients k_1, k_2. The control set is $A := \mathbf{R}$, and the set of admisible controls in each state x is the interval

$$A(x) := [-k_1|x|/k_2, k_1|x|/k_2]. \tag{50}$$

The disturbances z_t consists of i.i.d. random variables with values in $Z := \mathbf{R}$, zero mean and finite variance, that is,

$$E(z_t) = 0, \quad \sigma^2 := E(z_t^2) < \infty. \tag{51}$$

To complete the description of our control model we introduce the quadratic cost-per-stage function

$$C(x, a) := c_1 x^2 + c_2 a^2 \quad \forall (x, a) \in \mathbf{K}, \tag{52}$$

with positive coefficients c_1, c_2. We also define

$$W(x) := \exp[\gamma|x|] \quad \text{for all} \quad x \in \mathbf{X}, \tag{53}$$

with $\gamma \geq 4$. clearly, Assumption 3.4 holds. Moreover, let $\hat{s} > 0$ be such that

$$\gamma\hat{s} < \log(\gamma/2 + 1),$$

which implies

$$\beta := \frac{2}{\gamma}(\exp[\gamma\hat{s}] - 1) < 1. \tag{54}$$

Throughout the rest of this section, we suppose the following Assumptions taken from [6, Section 5]:

Assumption 5.1 $0 < k_1 < 1/2$.

Assumption 5.2 *The i.i.d. disturbances* z_t *have a common density* g*, which is a continuous bounded function supported on the interval* $S := [-\hat{s}, \hat{s}]$*. Moreover, there exists a positive number* ε *such that* $g(s) \geq \varepsilon$ *for all* $s \in S$*.*

These assumptions, 5.1 and 5.2, imply that Assumptions 3.1 and 3.2 hold (see, for instance,[6, Propositions 6, 23 and 24]).

On the other hand, in [6] it is proved that there exists a unique canonical policy given by

$$f_*(x) = -f_0 x, \quad \forall x \in \mathbf{X}, \tag{55}$$

satisfying (4) and (5), with

$$f_0 := \frac{v_0 k_1 k_2}{c_2 + v_0 k_2^2}$$

and v_0 is the unique positive solution to the quadratic (so-called Riccati) equation

$$k_2^2 v_0^2 + (c_2 - c_1 k_2^2 - c_2 k_1^2) v_0 - c_1 c_2 = 0.$$

In this case, the corresponding function $h_1(\cdot)$ is given by

$$h_1(x) = v_0 x^2 \quad \forall x \in \mathbf{X}, \tag{56}$$

and the optimal value is

$$\rho_* = v_0 \sigma^2, \tag{57}$$

where σ as in (51). Thus (ρ_*, h_1, f_*) is a canonical triplet for our linear quadratic Markov control model.

Since f_* in (55) is the unique canonical policy, by Proposition 3.6 we have that this policy also minimizes the limit average variance. In particular, the optimal value for the variance is

$$\sigma_*^2 = V(f_*, x) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E_x^{f_*} \Lambda_{f_*}(x_t), \tag{58}$$

We next calculate the limit in (58) and find the value of the optimal variance. To this end, let $\widehat{k} := k_1 - k_2 f_0$, $B := \int_{\mathbf{R}} z^3 g(z)\, dz$ and $D := \int_{\mathbf{R}} z^4 g(z)\, dz$. Then by (16), (55) and (56), we have

$$\Lambda_{f_*}(x_t) = v_0^2 \Big[4\widehat{k}^2 \sigma^2 E_x^{f_*}(x_t^2) + 4\widehat{k} B E_x^{f_*}(x_t) + D - \sigma^4 \Big], \tag{59}$$

Replacing a_t in (49) with $a_t := f_*(x_t) = -f_0 x_t$, we obtain

$$x_t = (k_1 - k_2 f_0)x_{t-1} + z_{t-1} = \widehat{k}x_{t-1} + z_{t-1} \quad \forall t = 1, 2, \cdots.$$

By (50) and Assumption 5.1, we can check that $|\widehat{k}| < 1$.

By an induction procedure, for all $t = 1, 2, \cdots$,

$$x_t = \widehat{k}^t x_0 + \sum_{j=0}^{t-1} \widehat{k}^j z_{t-1-j}.$$

From this relation, we obtain

$$E_x^{f_*}(x_t) = \widehat{k}^t x, \tag{60}$$

and

$$E_x^{f_*}(x_t^2) = \widehat{k}^{2t}x^2 + \sigma^2(1 - \widehat{k}^{2t})/(1 - \widehat{k}^2). \tag{61}$$

The relations (60) and (61) imply the limits

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E_x^{f_*}(x_t) = 0 \text{ and } \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E_x^{f_*}(x_t^2) = \sigma^2/(1 - \widehat{k}^2). \tag{62}$$

Hence, by (59) and (62) we obtain

$$\begin{aligned} \sigma_*^2 &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E_x^{f_*} \Lambda_{f_*}(x_t) \\ &= v_0^2 \Big[\frac{5\widehat{k}^2 - 1}{1 - \widehat{k}^2} \sigma^4 + \int_{\mathbf{R}} z^4 g(z)\, dz \Big] \geq 0. \end{aligned} \tag{63}$$

Finally, by Theorem 3.7 and considering (57), we obtain that for every initial state $x \in \mathbf{X}$, as $n \to \infty$, the distribution of the cost

$$\frac{\sum_{t=0}^{n-1} C_{f_*}(x_t) - n v_0 \sigma^2}{\sqrt{n}}$$

has an asymptotic normal distribution $N(0, \sigma_*^2)$ with σ_*^2 as in (63).

By (5), we obtain $v_0(1 - \widehat{k}^2) = c_1 + c_2 f_0^2$. Hence, $C_{f_*}(x) = (c_1 + c_2 f_0^2)x^2 = v_0(1 - \widehat{k}^2)x^2$ for all x. This implies that for every initial state x, as $n \to \infty$,

$$\frac{\sum_{t=0}^{n-1} x_t^2 - n\sigma^2/(1 - \widehat{k}^2)}{\sqrt{n}}$$

has asymptotic normal distribution $N(0, s^2)$, where

$$s^2 = \Big[\frac{5\widehat{k}^2 - 1}{1 - \widehat{k}^2}\sigma^4 + \int_{\mathbf{R}} z^4 g(z)\ dz\Big]\Big/(1 - \widehat{k}^2)^2.$$

Acknowledgement

The author wishes to thank Professor Onésimo Hernández-Lerma for his valuable comments and suggestions.

Armando F. Mendoza-Pérez
Universidad Politécnica
de Chiapas,
Calle Eduardo J.Selvas S/N,
Tuxtla Gutiérrez, Chiapas.
mepa680127@hotmail.com

References

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

[2] Hernández-Lerma O. and Lasserre J.B., *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, (1996).

[3] Hernández-Lerma O. and Lasserre J.B ., *Further Topics on Discrete-time Markov Control Processes*, Springer-Verlag, New York, (1999).

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

[5] Hernández-Lerma O., Vega-Amaya O. and Carrasco G., *Sample-path optimality and variance-minimization of average cost Markov control processes*, SIAM J. Control Optim., 38(1) (1999), 79-93.

[6] Hilgert N. and Hernández-Lerma O., *Bias optimality versus strong 0-discount optimality in Markov control processes with unbounded costs*, Acta Appl. Math. 77 (2003), 215-235.

[7] Mandl P., *On the variance in controlled Markov chains*, Kybernetika (Prague), 7 (1971), 1-12.

[8] Mandl P., *On the asymptotic normality of the reward in a controlled Markov chain*, Colloquia Mathematica Societatis János Bolyai, 9. European Meeting of Statisticians, Budapest (Hungary), (1972).

[9] Mandl P., *A connection between controlled Markov chains and martingales*, Kybernetika (Prague), 9 (1973), 237-241.

[10] Mandl P., *Estimation and control in Markov chains*, Adv. Appl. Probab., 6 (1974), 40-60.

[11] Prieto-Rumeau T. and Hernández-Lerma O., *Variance minimization and the overtaking optimality approach to continuous–time controlled Markov chains*, To appear in Math. Meth. Oper. Res.

[12] Puterman M.L., *Markov Decision Process*, Wiley, New York, (1994).

[13] Vega-Amaya O., *Markov control processes in Borel spaces: Undiscounted criteria*, Doctoral thesis, UAM-Iztapalapa, México, 1998 (in Spanish).

[14] Yushkevich A.A., *On a class of strategies in general Markov decision models*, Theory Probab. Appl., 18 (1973), 777-779.

[15] Zhu Q.X. and Guo X.P., *Markov decision processes with variance minimization: A new condition and approach*, Stoch. Anal. Appl., 25 (2007), 577-592.