Petr Mandl; Ma.Rosario Romera Ayllón
On adaptive control of Markov processes

Kybernetika, Vol. 23 (1987), No. 2, 89--103

Persistent URL: http://dml.cz/dmlcz/124869

KYBERNETIKA — VOLUME 23 (1987), NUMBER 2

ON ADAPTIVE CONTROL OF MARKOV PROCESSES

PETR MANDL, Mª ROSARIO ROMERA AYLLÓN

We consider a countable controlled Markov process. The control policy is the adaptive one. The differential generator is represented by the transition rate matrix. The case of infinite planning horizon and average reward per unit time as criterion is considered.

Sufficient conditions to ensure the existence of an optimal stationary strategy are given. The dependence of the asymptotic behaviour of the criterion function on the asymptotic behaviour of the control is also investigated. The paper is an extension to the continuous time case of the results in [3].

1. INTRODUCTION

We consider a system S with countable state space I. The process $X = \{X_t, t \geqq 0\}$ describes the trajectory of the system. The control process is $Z = \{Z_t, t \geqq 0\}$ ranging in a compact metric space $\mathscr{Z}$. The transition rate matrix Q is conservative, and the transition rates are continuous functions of z,

$$Q = \|q(i, j; z)\|_{i,j \in I}, \; z \in \mathscr{Z}, \quad q(i, z) = -q(i, i; z), \; i \in I, \; z \in \mathscr{Z}.$$

The criterion function is interpreted as the reward from S; up to time T it is given by

$$R_T = \int_0^T r(X_t, Z_t) \, \mathrm{d}t,$$

where $r(i, z)$ is a continuous function in z, $i \in I$, $z \in \mathscr{Z}$.

Admissible controls are defined following [4].

We consider the non-decreasing system of σ-algebras $\mathscr{F} = \{\mathscr{F}_t, t \geqq 0\}$ generated by the process $X = \{X_t, t \geqq 0\}$, which takes values in I, having piecewise constant right-continuous trajectories,

$$\mathscr{F}_t = \sigma a\{X_s, s \leqq t; \text{events of probability } 0\}.$$

The control process $Z = \{Z_t, t \geqq 0\}$ with values in $\mathscr{Z}$ is right-continuous and adapted to $\mathscr{F}$.

We say that X is the evolution of S under admissible control Z if

$$^{i}M_t = \chi\{X_t = i\} - \int_0^t q(X_s, i; Z_s)\,\mathrm{d}s\,, \quad t \geqq 0\,, \quad i \in I\,,$$

are martingales with respect to $\mathscr{F}$.

In the general case the control strategies considered are the non-anticipative ones. For stationary controls the control parameter at time t equals

$$Z_t = (\sigma)_{X_t}\,, \quad t \geqq 0\,,$$

where $\sigma \in \mathscr{Z} \times \mathscr{Z} \times \mathscr{Z} \ldots = \mathscr{Z}^{\infty}$.

The transition rate matrix under a stationary strategy has the expression

$$Q_\sigma = \|q(i, k; (\sigma)_i)\|_{i,k\in I} = \|q_\sigma(i, k)\|_{i,k\in I}\,,$$

and the reward up to time T it is given by

$$R_T = \int_0^T r(X_t, (\sigma)_{X_t})\,\mathrm{d}t = \int_0^T (r_\sigma)_{X_t}\,\mathrm{d}t\,.$$

To each stationary strategy σ we associate the mean reward defined as

$$\lim_{T\to\infty} T^{-1} R_T = \mu_\sigma\,,$$

provided that the limit exists and is constant with probability one.

We take optimality in the following sense; $\hat{\sigma}$ is an optimal stationary policy, if

$$\mu_{\hat{\sigma}} = \hat{\mu} = \sup_{\sigma\in\mathscr{Z}^{\infty}} \mu_\sigma\,.$$

Trajectory of S Corresponding to a Stationary Policy

Without loss of generality we select the state 1 to be the exceptional one. Following the idea expressed by Hordijk ([2]) in order to obtain sufficient conditions for the existence of an optimal stationary strategy, we introduce the $\tilde{Q}_\sigma$ matrix

$$\tilde{q}_\sigma(i, j) = \begin{cases} q_\sigma(i, j), & i, j \in I\,, \quad i \neq 1\,, \\ q_\sigma(1, j)\,\delta_{1j}\,, & j \in I\,, \end{cases}$$

which is the restriction of Q_σ, obtained by replacing every element in the first column with exception of the diagonal one by 0. δ is the Kronecker delta function.

The matrix $\tilde{Q}_\sigma$ then corresponds to the situation when the trajectory of S vanishes at the jump into state 1.

Let $\tilde{P}_\sigma(t)$, $t \geqq 0$, be the minimal solution of the backward equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_\sigma(t) = \tilde{Q}_\sigma\,\tilde{P}_\sigma(t)$$

$$\tilde{P}_\sigma(0) = \boldsymbol{I}\,(\text{unit matrix})\,.$$

With a stationary policy σ we associate $X = \{X_t, t \geqq 0\}$ describing the evolution

of S under σ as follows. From the initial state $X_0 = i$ the process X follows the transition law $\tilde{P}_\sigma$ up to the first vanishing of the trajectory. The trajectory is right continuous. After the vanishing, the process starts afresh in state 1, and follows $\tilde{P}_\sigma$ until it vanishes again and so on.

The jump-matrix corresponding to the Markov chain embedded in process X is

$$\tilde{P}_\sigma = \|\tilde{P}_\sigma(i,j)\|_{i,j\in I},$$

where

$$\tilde{p}_\sigma(i,j) = \begin{cases} 0, & i = j, \\ \dfrac{\tilde{q}_\sigma(i,j)}{\tilde{q}_\sigma(i)}, & i \neq j, \end{cases} = \frac{\tilde{q}_\sigma(i,j)}{\tilde{q}_\sigma(i)}(1 - \delta_{ij}), \quad i,j \in I.$$

The semi-Markovian viewpoint gives the following expressions

$$\begin{aligned} \int_0^\infty \tilde{P}_\sigma(s)\, e\, \mathrm{d}s &= \sum_{n=0}^{\infty} \tilde{P}_\sigma^n \frac{e}{q_\sigma}, \\ \int_0^\infty \tilde{P}_\sigma(s)\, r_\sigma\, \mathrm{d}s &= \sum_{n=0}^{\infty} \tilde{P}_\sigma^n \frac{r_\sigma}{q_\sigma}, \end{aligned} \tag{1}$$

provided that the series converge absolutely. We set $e = (1, 1, \ldots, 1)'$.

If v is a vector, then $(|v|)_i = |(v)_i|$, $(v^2)_i = (v)_i^2$. Analogously

$$\left(\frac{v_1}{v_2}\right)_i = \frac{(v_1)_i}{(v_2)_i} \quad \text{for vectors } v_1 \text{ and } v_2.$$

2. EXISTENCE OF AN OPTIMAL STATIONARY STRATEGY

We consider optimality in the maximum average reward per unit time sense; $\hat{\sigma} \in \mathscr{Z}^\infty$ is optimal if

$$\mu_{\hat{\sigma}} = \sup_{\sigma \in \mathscr{Z}^\infty} \mu_\sigma = \hat{\mu}.$$

In order to ensure the stability we introduce a Liapunov-type condition, which is similar to that introduced by Hordijk in [2].

Hypothesis H 1. There exists a vector $y_1 \geqq 0$ such that

(i) $|r_\sigma| + e + \tilde{Q}_\sigma y_1 \leqq 0, \quad \sigma \in \mathscr{Z}^\infty,$

(ii) $\lim_{N\to\infty} \tilde{P}_\sigma^N y_1 = 0, \quad \sigma \in \mathscr{Z}^\infty,$

(iii) $\lim_{\sigma\to\sigma_0} \tilde{Q}_\sigma y_1 = \tilde{Q}_{\sigma_0} y_1, \quad \sigma, \sigma_0 \in \mathscr{Z}^\infty.$

The convergence considered in $\mathscr{Z}^\infty$ is the component-wise one given by the metric in the compact space $\mathscr{Z}^\infty$.

Example. Conditions H 1 do not exclude the runs of the trajectory to infinity.

This is seen on the divergent pure birth process with rates

$$q(j, j+1) = j^2 = -q(j, j), \quad j = 1, 2, \ldots$$

The jth inequality in (i) is

$$(|r_\sigma|)_j + 1 - j^2(y_1)_j + j^2(y_1)_{j+1} \leqq 0. \tag{2}$$

Let

$$(|r_\sigma|)_j + 1 \leqq \text{const.}, \quad j = 1, 2 \ldots$$

Hence (2) holds for

$$(y_1)_j = 2\,\text{const.}/j, \quad j = 1, 2, \ldots$$

Lemma 1. The Liapunov type condition (i) ensures the existence of μ_σ for all $\sigma \in \mathscr{Z}^\infty$.

Proof. Having in consideration that one method to construct the $\tilde{P}_\sigma$ minimal backward Kolmogorov solution is the truncation procedure, we introduce the ${}^n\tilde{P}_\sigma$ solution of the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\,{}^n\tilde{P}_\sigma(t) = {}^n\tilde{Q}_\sigma\,{}^n\tilde{P}_\sigma(t), \quad {}^n\tilde{P}_\sigma(0) = \boldsymbol{I},$$

where ${}^n\tilde{Q}_\sigma$ is the $n \times n$ truncation of $\tilde{Q}_\sigma$.

We also consider truncated vectors ${}^n e$, ${}^n y_1$. Then we have

$${}^n e + {}^n\tilde{Q}_\sigma\,{}^n y_1 \leqq 0,$$

$${}^n\tilde{P}_\sigma(s)\,{}^n e + {}^n\tilde{P}_\sigma(s)\,{}^n\tilde{Q}_\sigma\,{}^n y_1 \leqq 0,$$

$${}^n\tilde{P}_\sigma(s)\,{}^n e + \frac{\mathrm{d}}{\mathrm{d}s}\,{}^n\tilde{P}_\sigma(s)\,{}^n y_1 \leqq 0,$$

and hence

$$\int_0^t {}^n\tilde{P}_\sigma(s)\,{}^n e\,\mathrm{d}s + {}^n\tilde{P}_\sigma(t)\,{}^n y_1 \leqq {}^n\tilde{P}_\sigma(0)\,{}^n y_1 = {}^n y_1.$$

Letting $n \to \infty$ and then $t \to \infty$, we obtain

$$\int_0^\infty \tilde{P}_\sigma(s)\, e\,\mathrm{d}s \leqq y_1. \tag{3}$$

Similarly

$$\int_0^\infty \tilde{P}_\sigma(s)\,|r_\sigma|\,\mathrm{d}s \leqq y_1. \tag{4}$$

From (3) it follows that under σ state 1 is positively recurrent. From (4) it follows that the expected reward earned until the jump into state 1 is finite. Hence, the strong law of large numbers yields

$$\lim_{T\to\infty} \frac{1}{T} R_T = \frac{\left(\int_0^\infty \tilde{P}_\sigma(s)\, r_\sigma\,\mathrm{d}s\right)_1}{\left(\int_0^\infty \tilde{P}_\sigma(s)\, e\,\mathrm{d}s\right)_1} = \mu_\sigma \quad \text{a.s.} \qquad \square$$

If we make the additional assumption

$$q(i, 1; z) \leqq \text{const.}, \quad z \in \mathscr{Z}, \quad i = 2, 3, \ldots,$$

it follows with regard to (1) and to (4) that $\hat{\mu}$ is finite.

We also denote $\bar{q} = \sup_{\substack{i \in I \\ z \in \mathscr{Z}}} q(i, 1; z)$.

Theorem 1. Let H 1 hold. Then there exists an optimal stationary policy.

In virtue of (1) the proof can proceed as in [2].

Next define the potential

$$w = \sup_{\sigma_0 \sigma_1 \ldots} \sum_{n=0}^{\infty} \tilde{P}_{\sigma_0} \tilde{P}_{\sigma_1} \ldots \tilde{P}_{\sigma_{n-1}} \left(\frac{r_{\sigma_n}}{q_{\sigma_n}} - \hat{\mu} \frac{e}{q_{\sigma_n}} \right),$$

which represents the optimal expected reward if $r(i, z) - \hat{\mu}$ is taken as reward function.

We have the optimality equations

$$w = \sup_{\sigma} \left\{ \frac{r_\sigma}{q_\sigma} - \frac{\hat{\mu} e}{q_\sigma} + \tilde{P}_\sigma w \right\}, \tag{5}$$

$$w = \frac{r_{\hat{\sigma}}}{q_{\hat{\sigma}}} - \frac{\hat{\mu} e}{q_{\hat{\sigma}}} + \tilde{P}_{\hat{\sigma}} w, \tag{6}$$

which are similar to the discrete case optimality equations obtained in [3].

(5) and (6) are equivalent to

$$0 = \sup_{\sigma} \{ r_\sigma - \hat{\mu} e + Q_\sigma w \}, \tag{7}$$

$$0 = r_{\hat{\sigma}} - \hat{\mu} e + Q_{\hat{\sigma}} w. \tag{8}$$

We can conclude that $(\hat{\mu}, \hat{\sigma})$ is the solution of the optimality equations. We also have

$$0 = r_{\bar{\sigma}} - \mu_{\bar{\sigma}} e + Q_{\bar{\sigma}} w_{\bar{\sigma}} \quad \text{for} \quad \bar{\sigma} \in \mathscr{Z}^\infty.$$

3. ASYMPTOTIC BEHAVIOUR OF THE CRITERION FUNCTIONAL

In this section we consider the initial state X_0 fixed.

Lemma 2. Let $0 \leqq c(i, z)$, $i \in I$, $z \in \mathscr{Z}$, be continuous function of z. Let

$$c_\sigma + \tilde{Q}_\sigma y \leqq 0, \quad \sigma \in \mathscr{Z}^\infty, \tag{9}$$

with $y \geqq 0$.

Then, under admissible control Z for $T \geqq 0$

$$\mathsf{E} \int_0^T c(X_t, Z_t) \, \mathrm{d}t + \mathsf{E}(y)_{X_T} \leqq (y)_{X_0} + (y)_1 \, \bar{q} T. \tag{10}$$

Proof. Let $n \geqq X_0$. Denote $\tau_n = \inf\{t, X_t > n\}$. Then

$$\{{}^iM_{t\wedge\tau_n},\ t \geqq 0\}, \quad i \in I, \quad \text{where} \quad t \wedge \tau_n = \min(t, \tau_n),$$

are martingales, and

$$\sum_{i=1}^{m} (y)_i\, {}^iM_{t\wedge\tau_n}, \quad t \geqq 0,$$

are also martingales for each m.

Let $T \geqq 0$. For $m > X_0$ we have

$$(y)_{X_0} = \mathsf{E}\sum_{i=1}^{m} (y)_i\, {}^iM_{T\wedge\tau_n} =$$

$$= \mathsf{E}\chi\{X_{T\wedge\tau_n} \leqq m\}\,(y)_{X_{T\wedge\tau_n}} - \mathsf{E}\int_0^{T\wedge\tau_n} \sum_{i=1}^{m} q(X_t, i; Z_t)\,(y)_i\,\mathrm{d}t\,.$$

Letting $m \to \infty$ we obtain

$$(y)_{X_0} = \mathsf{E}(y)_{X_{T\wedge\tau_n}} - \mathsf{E}\int_0^{T\wedge\tau_n} \sum_i q(X_t, i; Z_t)\,(y)_i\,\mathrm{d}t\,. \tag{11}$$

Note that the last expectation is finite, since from (9) follows

$$\left|\sum_i q(X_t, i; Z_t)\,(y)_i\right| \leqq q(X_t, Z_t)\,(y)_{X_t} + \left|q(X_t, 1; Z_t)\right|(y)_1\,,$$

and in the integrand is $X_t \in \{1, \ldots, n\}$ for $t \in [0, T \wedge \tau_n]$. From (11) and (9)

$$(y)_{X_0} - \mathsf{E}\int_0^{T\wedge\tau_n} c(X_t, Z_t)\,\mathrm{d}t = \mathsf{E}(y)_{X_{T\wedge\tau_n}} - \mathsf{E}\int_0^{T\wedge\tau_n} \chi\{X_t \neq 1\}\, q(X_t, 1; Z_t)\,(y)_1\,\mathrm{d}t -$$

$$- \mathsf{E}\int_0^{T\wedge\tau_n} (c(X_t, Z_t) + \sum_i \tilde{q}(X_t, i; Z_t)\,(y)_i)\,\mathrm{d}t \geqq \mathsf{E}\chi\{\tau_n > T\}\,(y)_{X_T} - \bar{q}(y)_1 \mathsf{E}(T \wedge \tau_n)\,.$$

Letting $n \to \infty$ (10) is obtained. □

Law of Large Numbers

In the following we fix $\bar{\sigma} \in \mathscr{Z}^\infty$. To $\bar{\sigma}$ we can associate the scalar-vector couple $(\mu_{\bar{\sigma}}, w_{\bar{\sigma}})$ satisfying

$$0 = r_{\bar{\sigma}} - \mu_{\bar{\sigma}} e + Q_{\bar{\sigma}} w_{\bar{\sigma}}\,.$$

We shall write briefly μ, w.

We define

$$\varphi(i, z) = r(i, z) - \mu + \sum_j q(i, j; z)\,(w)_j\,, \quad i \in I, \quad z \in \mathscr{Z}\,,$$

which is a continuous function in z. $\varphi(X_t, Z_t)$ is a measure of difference between the actual parameter value Z_t and the value $(\bar{\sigma})_{X_t}$ corresponding to the stationary strategy $\bar{\sigma}$.

We introduce the process

$$M_T = R_T - T\mu + (w)_{X_T} - (w)_{X_0} - \int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t\,, \quad T \geqq 0\,, \tag{12}$$

and the functions for $m = 2, 3, \ldots$

$$r_m(i, z) = \sum_{j \neq i} q(i, j; z)\,((w)_j - (w)_i)^m ,$$

$$h_m(i, z) = \sum_{j \neq i} q(i, j; z)\,((y_1)_j^m + (y_1)_i^m) , \quad i \in I , \quad z \in \mathscr{Z} .$$

In order to prove that the Law of Large Numbers holds for R_T, we need to make a stronger hypothesis.

Hypothesis H 2. There exists a $y_2 \geqq y_1^2$ such that

$$h_{2\sigma} + \tilde{Q}_\sigma y_2 \leqq 0 , \quad \sigma \in \mathscr{Z}^\infty .$$

Since $|w| \leqq$ const. y_1 we have,

$$|r_m(i, z)| \leqq \text{const. } h_m(i, z) .$$

Lemma 3. Let H 1 and H 2 hold. Then for any admissible Z,

(i) $M = \{M_T, T \geqq 0\}$ is a martingale with respect to $\mathscr{F}$.

(ii) $\mathsf{E}(M_T - M_S)^2 = \mathsf{E} \int_S^T r_2(X_t, Z_t)\,\mathrm{d}t , \quad 0 \leqq S \leqq T.$

(iii) $\mathsf{E}M_T^2 + \mathsf{E}(y_2)_{X_T} \leqq (y_2)_{X_0} + (y_2)_1\,\bar{q}T , \quad T \geqq 0.$

Proof. Detailed proof is given in [5], Lemma III.2.1. From Lemma 2 one obtains

$$\mathsf{E} \int_0^{T \wedge \tau_n} r_2(X_t, Z_t)\,\mathrm{d}t + \mathsf{E}(y_2)_{X_{T \wedge \tau_n}} \leqq (y_2)_{X_0} + (y_2)_1\,\bar{q}T , \quad T \geqq 0 ,$$

where $\tau_n = \inf\{t, X_t > n\}$.

First it is shown that $\{M_{T \wedge \tau_n}, T \geqq 0\}$, $n = 1, 2, \ldots$, are martingales and that

$$\mathsf{E}(M_{T \wedge \tau_n})^2 = \mathsf{E} \int_0^{T \wedge \tau_n} r_2\,\mathrm{d}t .$$

Letting $n \to \infty$ (i), (ii), (iii) follow.

From this lemma we can get the following results.

Theorem 1. Let H 1, H 2 hold. Then

$$\lim_{T \to \infty} T^{-1} R_T = \mu$$

in probability, in first order mean, in second order mean if and only if

$$\lim_{T \to \infty} T^{-1} \int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t = 0 ,$$

respectively in probability, in first order mean and in second order mean.

Theorem 2. Let H 1, H 2 hold. Then

$$\lim_{T \to \infty} T^{-1} M_T = 0 \quad \text{a.s.} \tag{13}$$

Theorem 3. Let H 1, H 2 hold. Then under any admissible control

$$\lim_{T\to\infty} \mathsf{P}(T^{-1}R_T \geqq \hat{\mu} + \varepsilon) = 0 \quad \text{for} \quad \varepsilon > 0\,,$$

and

$$\overline{\lim_{T\to\infty}}\, T^{-1}\,\mathsf{E}R_T \leqq \hat{\mu}\,.$$

The proof of these theorems can proceed as in [5].

In order to obtain the main result of this subsection, we introduce the vectors e_k

$$e_k\colon (e_k)_i = 0\,, \quad i < k\,, \quad (e_k)_i = 1\,, \quad i \geqq k\,,$$

and set

$$\varepsilon_\sigma(k) = \left(\int_0^\infty \tilde{P}_\sigma'(s)\, e_k\, \mathrm{d}s\right)_1 \Big/ \left(\int_0^\infty \tilde{P}_\sigma'(s)\, e\, \mathrm{d}s\right)_1\,, \quad \sigma \in \mathscr{Z}^\infty\,,$$

$$\hat{\varepsilon}(k) = \sup_{\sigma\in\mathscr{Z}^\infty} \varepsilon_\sigma(k)\,.$$

Following [3], Lemma 10, we can derive

$$\lim_{k\to\infty} \hat{\varepsilon}(k) = 0\,.$$

Since H 1 implies

$$e_k + e + 2\tilde{Q}_\sigma y_1 \leqq 0\,, \quad \sigma \in \mathscr{Z}^\infty\,,$$

we can use Theorem 3 to conclude that under H 1, H 2

$$\overline{\lim_{T\to\infty}}\, T^{-1}\,\mathsf{E}\int_0^T (e_k)_{X_t}\,\mathrm{d}t = \overline{\lim_{T\to\infty}}\, T^{-1}\int_0^T \mathsf{P}(X_t \geqq k)\,\mathrm{d}t \leqq \hat{\varepsilon}(k)\,,$$

for Z admissible.

The analogy between $\varepsilon_\sigma(k)$, $\hat{\varepsilon}(k)$ and the quantities μ_σ, $\hat{\mu}$ is obvious.

Hypothesis H 3. There exists a $y_3 \geqq 0$ such that

$$(q_\sigma y_1)^2 + \tilde{Q}_\sigma y_3 \leqq 0\,, \quad \sigma \in \mathscr{Z}^\infty\,.$$

Theorem 4. Let H 1, H 2, H 3 hold. If the admissible control Z is such that

$$\lim_{t\to\infty} \varrho(Z_t, (\bar{\sigma})_{X_t}) = 0 \quad \text{in probability}\,, \tag{14}$$

(ϱ is the distance), then

$$\lim_{T\to\infty} \mathsf{E}\left|T^{-1}R_T - \mu\right| = 0\,.$$

Proof. Let (14) hold. According to Theorem 1 we have to establish

$$\lim_{T\to\infty} T^{-1}\,\mathsf{E}\left|\int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t\right| = 0\,.$$

H 1 implies

$$|r(i, z)| \leqq q(i, z)\,(y_1)_i\,,$$

$$\Big|\sum_j q(i, j, z)\,(w)_j\Big| \leqq \text{const.}\Big|\sum_{j \neq i} q(i, j; z)\,(y_1)_j + q(i, z)\,(y_1)_i\Big| \leqq$$

$$\leqq \text{const.}\,(q(i, z)\,(y_1)_i + \bar{q}(y_1)_1)\,.$$

Hence we conclude that

$$|\varphi(i, z)|^2 \leqq C[1 + (q(i, z)\,(y_1)_i)^2]\,, \tag{15}$$

where C is a constant. φ is a continuous function in z, in virtue of H 1, (iii).

Take $L > 0$, $k > 0$, k integer. It holds

$$T^{-1}\,\mathsf{E}\left|\int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t\right| \leqq T^{-1}\int_0^T \mathsf{E}|\chi\{X_t < k\}\,\varphi(X_t, Z_t)|\,\mathrm{d}t + \tag{16}$$

$$+\,T^{-1}\,L\int_0^T \mathsf{P}(X_t \geqq k)\,\mathrm{d}t + T^{-1}\int_0^T \mathsf{E}|\chi\{\varphi(X_t, Z_t) > L\}\,\varphi(X_t, Z_t)|\,\mathrm{d}t\,.$$

We have

$$\lim_{t\to\infty} \mathsf{E}|\chi\{X_t < k\}\,\varphi(X_t, Z_t)| = 0\,,$$

since $\chi\{X_t < k\}\,\varphi(X_t, Z_t)$ is bounded and tends to 0 in probability as $t \to \infty$, by (14) and by $\varphi(X_t, (\bar{\sigma})_{X_t}) = 0$.

The second term on the right in (15) is estimated from the expression obtained before,

$$\overline{\lim_{T\to\infty}}\, T^{-1}\int_0^T \mathsf{P}(X_t \geqq k)\,\mathrm{d}t \leqq \hat{\varepsilon}(k)\,.$$

As far as the third term, we can use H 3, (15) and Lemma 2 to derive

$$\int_0^T \mathsf{E}\varphi^2(X_t, Z_t)\,\mathrm{d}t \leqq C\left(T + \int_0^T (q(X_t, Z_t)\,(y_1)_{X_t})^2\,\mathrm{d}t\right) \leqq C(T + (y_3)_{X_0} + (y_3)_1\,\bar{q}T)\,.$$

Since $\mathsf{E}|\chi\{\varphi(X_t, Z_t) > L\}\,\varphi(X_t, Z_t)| \leqq L^{-1}\,\mathsf{E}\varphi^2(X_t, Z_t)$, we have that

$$\overline{\lim_{T\to\infty}}\, T^{-1}\int_0^T \mathsf{E}|\{\varphi(X_t, Z_t) > L\}\,\varphi(X_t, Z_t)|\,\mathrm{d}t \leqq L^{-1}C(1 + (y_3)_1\,\bar{q})\,.$$

Consequently,

$$\overline{\lim_{T\to\infty}}\, T^{-1}\left|\int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t\right| \leqq L\hat{\varepsilon}(k) + L^{-1}C(1 + (y_3)_1\,\bar{q})\,.$$

The right-hand side can be made arbitrarily small by taking L and then k sufficiently large. □

The results aforementioned allow to say that the validity of the Law of Large Numbers for R_T holds if the difference (measured in distance) between the actual parameter control value Z_t and that corresponding to the stationary strategy $(\bar{\sigma})_{X_t}$ tends to zero if t goes to infinity.

Central Limit Theorem

Our next purpose is to prove the Central Limit Theorem for R_T.

We introduce hypotheses H* 1, H* 2 and H* 3. They are analogous to H 1, H 2 and H 3 with r_σ replaced by $h_{2\sigma}$ (or $r(i, z)$ replaced by $h_2(i, z)$).

Set $Y_n = M_{n+1} - M_n$, $n = 0, 1 \ldots$, defined from the martingale (12).

We consider a fixed $\bar{\sigma} \in \mathscr{Z}^\infty$ and denote $\mu_2 = \lim_{N\to\infty} N^{-1} \int_0^N r_2(X_t, Z_t)\, dt$.

From martingale theory ([1]) the following result is known.

Proposition 1. Let $\{Y_n, n = 0, 1 \ldots\}$ be the martingale differences. Further let

$$\lim_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} \mathsf{E}\{Y_n^2 \mid \mathscr{F}_n\} = v \quad \text{in probability}$$

where v is a constant, and for each $\varepsilon > 0$ let the Lindeberg condition hold

$$\lim_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} \mathsf{E} Y_n^2 \chi\{|Y_n| \geqq \varepsilon \sqrt{N}\} = 0\,.$$

Then

$$M_N/\sqrt{N} \text{ as } N \to \infty \text{ has asymptotically normal distribution } N(0, v)\,.$$

Extending the argument of Lemma 3 we can prove

$$\mathsf{E}\{Y_n^2 \mid \mathscr{F}_n\} = \mathsf{E}\left\{\int_n^{n+1} r_2(X_t, Z_t)\, dt \,\middle|\, \mathscr{F}_n\right\}.$$

The aim is to use the result from Proposition 1 so we need to verify its two conditions.

Concerning the first one we have the following. If H 1 and H 2 hold then

$$\lim_{N\to\infty} \left(N^{-1} \int_0^N r_2(X_t, Z_t)\, dt - N^{-1} \sum_{n=0}^{N-1} \mathsf{E}\{Y_n^2 \mid \mathscr{F}_n\}\right) = 0 \quad \text{a.s.}$$

To prove this, set

$$L_N = \int_0^N r_2(X_t, Z_t)\, dt - \sum_{n=0}^{N-1} \mathsf{E}\left\{\int_n^{n+1} r_2(X_t, Z_t)\, dt \,\middle|\, \mathscr{F}_n\right\}, \quad N = 1, 2, \ldots$$

Then $L = \{L_N, N = 1, 2, \ldots\}$ is a martingale, and we have that

$$\mathsf{E}(L_{N+1} - L_N)^2 = \mathsf{E}\left(\int_N^{N+1} r_2(X_t, Z_t)\, dt\right)^2 - \mathsf{E}\left(\mathsf{E}\left\{\int_N^{N+1} r_2(X_t, Z_t)\, dt \,\middle|\, \mathscr{F}_N\right\}\right)^2 \leqq$$

$$\leqq \mathsf{E}\left(\int_N^{N+1} r_2(X_t, Z_t)\, dt\right)^2 \leqq \mathsf{E} \int_N^{N+1} r_2^2(X_t, Z_t)\, dt\,, \quad N = 1, 2, \ldots$$

This, as in the proof of Theorem 2, implies

$$\sum_{n=1}^{\infty} \frac{\mathsf{E}(L_{n+1} - L_n)^2}{n^2} < \infty\,.$$

Consequently, L fulfills the Law of Large Numbers. Hence,

$$\lim_{N\to\infty}\left(N^{-1}\int_0^N r_2(X_t, Z_t)\,\mathrm{d}t - N^{-1}\sum_{n=0}^{N-1}\mathsf{E}\{Y_n^2 \mid \mathscr{F}_n\}\right) = 0 \quad \text{a.s.} \tag{17}$$

If H* 1, H* 2, H* 3 hold, we conclude from Theorem 4 that

$$\lim_{N\to\infty} N^{-1}\int_0^N r_2(X_t, Z_t)\,\mathrm{d}t = \mu_2 \quad \text{in probability}.$$

Hence, (17) implies

$$\lim_{N\to\infty} N^{-1}\sum_{n=0}^{N-1}\mathsf{E}\{Y_n^2 \mid \mathscr{F}_n\} = \mu_2 \quad \text{in probability}.$$

Concerning the second hypothesis of Proposition 1, namely the Lindeberg condition, we have to state an auxiliary result.

Lemma 4. Let H 1, H 2, H* 1, H* 2 and H* 3 hold. Then

$$\overline{\lim_{N\to\infty}}\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{E}Y_n^4 \leqq \text{const.} \tag{18}$$

under arbitrary admissible control Z.

The laborious proof can proceed as in [4] and is based on the relation

$$\mathsf{E}(M_T - M_S)^4 = \mathsf{E}\left(\int_S^T r_4(X_t, Z_t)\,\mathrm{d}t + 4\int_S^T (M_t - M_S)\, r_3(X_t, Z_t)\,\mathrm{d}t + 6\int_S^T (M_t - M_S)^2\, r_2(X_t, Z_t)\,\mathrm{d}t\right). \tag{19}$$

Next step deals with the expression

$$\begin{aligned}\sum_{n=0}^{N-1}\mathsf{E}(M_{n+1} - M_n)^4 &= \sum_{n=0}^{N-1}\mathsf{E}\int_0^1 r_4(X_{t+n}, Z_{t+n})\,\mathrm{d}t + \\ &+ 4\sum_{n=0}^{N-1}\mathsf{E}\int_0^1 (M_{t+n} - M_n)\, r_3(X_{t+n}, Z_{t+n})\,\mathrm{d}t + \\ &+ 6\sum_{n=0}^{N-1}\mathsf{E}\int_0^1 (M_{t+n} - M_n)^2\, r_2(X_{t+n}, Z_{t+n})\,\mathrm{d}t \leqq \mathsf{E}\int_0^N r_4(X_t, Z_t)\,\mathrm{d}t + \\ &+ 4\Big(\sum_{n=0}^{N-1}\mathsf{E}(M_{n+1} - M_n)^4\Big)^{1/4}\left(\mathsf{E}\int_0^N |r_3(X_t, Z_t)|^{4/3}\,\mathrm{d}t\right)^{3/4} + \\ &+ 6\Big(\sum_{n=0}^{N-1}\mathsf{E}(M_{n+1} - M_n)^4\Big)^{1/2}\left(\mathsf{E}\int_0^N r_2^2(X_t, Z_t)\,\mathrm{d}t\right)^{1/2}.\end{aligned}$$

This expression is obtained from (19), using the Hölder inequality and the fact that

$$\mathsf{E}(M_{t+n} - M_n)^4 \quad \text{is nondecreasing in } t.$$

It is not difficult to see that from here (18) follows, provided that

$$\mathsf{E}\int_0^N r_4(X_t, Z_t)\,\mathrm{d}t + \mathsf{E}\int_0^N |r_3(X_t, Z_t)|^{4/3}\,\mathrm{d}t + \mathsf{E}\int_0^N r_2^2(X_t, Z_t)\,\mathrm{d}t \leqq \text{const.}\, N.$$

From Lemma 4, the Lindeberg condition follows; for $\varepsilon > 0$

$$\overline{\lim_{N\to\infty}} N^{-1}\sum_{n=0}^{N-1}\mathsf{E}Y_n^2\chi\{|Y_n| \geqq \varepsilon\sqrt{N}\} \leqq \overline{\lim_{N\to\infty}} \varepsilon^{-2}N^{-2}\sum_{n=0}^{N-1}\mathsf{E}Y_n^4 = 0\,.$$

We can state now the main result in this subsection.

Theorem 5. Let H 1, H 2, H 3, H* 1, H* 2, H* 3 hold, and let Z be an admissible control such that

$$\lim_{t\to\infty} \varrho(Z_t, (\bar{\sigma})_{X_t}) = 0 \quad \text{in probability}\,, \tag{20}$$

$$\lim_{T\to\infty} \frac{1}{\sqrt{T}}\int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t = 0 \quad \text{in probability}\,.$$

Then $(R_T - T\mu)/\sqrt{T}$ has asymptotically normal distribution $N(0, \mu_2)$ as $T\to\infty$.

Proof. Set

$$\frac{M_T}{\sqrt{T}} = \frac{R_T - T\mu}{\sqrt{T}} - \frac{(w)_{X_T} - (w)_{X_0}}{\sqrt{T}} + \frac{1}{\sqrt{T}}\int_0^T \varphi(X_t, Z_t)\,\mathrm{d}t\,. \tag{21}$$

Theorem 4 gives the following relation

$$\lim_{N\to\infty} N^{-1}\int_0^N r_2(X_t, Z_t)\,\mathrm{d}t = \mu_2 \quad \text{in probability}\,.$$

Hence from (17)

$$\lim_{N\to\infty} N^{-1}\sum_{n=0}^{N-1}\mathsf{E}\{Y_n^2 \mid \mathscr{F}_n\} = \mu_2 \quad \text{in probability}\,.$$

We see that $M_N/\sqrt{N}$ is asymptotically $N(0, \mu_2)$ as $N\to\infty$. From Lemma 4 follows

$$\sum_{N=1}^{\infty}\frac{\mathsf{E}Y_N^4}{N^2} < \infty\,. \tag{22}$$

Further

$$\mathsf{E}\left(\frac{M_T}{\sqrt{[T]}} - \frac{M_{[T]}}{\sqrt{[T]}}\right)^4 \leqq \frac{\mathsf{E}Y_{[T]}^4}{[T]^2}\,.$$

In virtue of (22), the right term in this inequality converges to zero as T tends to ∞. We conclude that $M_T/\sqrt{[T]}$ and hence $M_T/\sqrt{T}$ is asymptotically $N(0, \mu_2)$ as T tends to infinity. The proof is accomplished by demonstrating the negligibility of

$$[(w)_{X_T} - (w)_{X_0}]/\sqrt{T} \quad \text{as } T \text{ tends to } \infty\,. \qquad \square$$

4. EXAMPLE

As an illustrative example we propose the consideration of a system described as a $M/M/1$ queue, in order to apply the results obtained before.

Let us suppose a known arrival rate given by q, and a known service rate given by z, which is going to be control parameter ranging in the interval $[z', z'']$, where $z' > q$.

The process X_t denotes the number of customers in the system at time $t > 0$. The cost functional is defined as follows

$$C_T = \int_0^T (X_t + KZ_t)\, \mathrm{d}t\,,$$

where K is a positive constant.

The Bellman equation from Section 2 for the minimum mean cost takes the expression

$$0 = \min_{z\in[z',z'']} [Kz - qw_0 + qw_1 - \hat{\mu}]\,,$$

$$0 = \min_{z\in[z',z'']} [i + Kz + zw_{i-1} - (z+q)\, w_i + qw_{i+1} - \hat{\mu}]\,, \quad i = 1, 2, \ldots$$

Then, minimizing the expression in brackets we obtain for z

$$z = z'' \quad \text{if} \quad K + w_{i-1} - w_i < 0\,,$$

$$z = z' \quad \text{if} \quad K + w_{i-1} - w_i > 0\,.$$

The optimal strategy of control takes the form

$$\sigma(i) = z'\,, \quad i = 1, 2 \ldots n\,,$$

$$\sigma(i) = z''\,, \quad i = n+1, n+2, \ldots,$$

which is a bang-bang control.

The average cost $\mu(n)$ corresponding to this strategy can be obtained in a direct way. The stationary distribution for the σ strategy is

$$p_i = g\left(\frac{q}{z'}\right)^i, \quad i = 0, 1, \ldots, n-1\,,$$

$$p_i = g\left(\frac{q}{z'}\right)^n \left(\frac{q}{z''}\right)^{i-n}, \quad i = n, \quad n+1, \ldots,$$

where

$$g = \left(\frac{1-u^m}{1-u} + u^n \frac{1}{1-v}\right)^{-1}, \quad u = \frac{q}{z'}, \quad v = \frac{q}{z''}.$$

According to the definition of the mean cost

$$\mu(n) = g\left\{u\,\frac{1-u^n}{(1-u)^2} - nu^n\left(\frac{1}{1-u} - \frac{1}{1-v}\right) + \right.$$

$$\left. + u^n\,\frac{v}{(1-v)^2} + K(z''-z')\,u^n\,\frac{v}{1-v}\right\} + Kz' .$$

The optimal value n verifies

$$\mu(\hat{n}) = \min_n \mu(n) = \hat{\mu} .$$

Then the optimal strategy of control $\hat{\sigma}$ fulfils

$$(\hat{\sigma})_i = z' , \quad i = 0, 1, \ldots, \hat{n} ,$$

$$(\hat{\sigma})_i = z'' , \quad i = \hat{n}+1, \ldots$$

The optimality equations obtained before can be expressed in the following way

$$0 = Kz' + q(w_1 - w_0) - \hat{\mu} ,$$

$$0 = i + Kz' - z'(w_i - w_{i-1}) + q(w_{i+1} - w_i) - \hat{\mu} , \quad i = 1, \ldots, \hat{n} ,$$

$$0 = i + Kz'' - z''(w_i - w_{i-1}) + q(w_{i+1} - w_i) - \hat{\mu} , \quad i = \hat{n}+1, \hat{n}+2 \ldots$$

w_i is a quadratic function for $i = \hat{n}+1, \hat{n}+2, \ldots$ The fact that $\hat{\sigma}$ is an optimal stationary strategy is expressed by the inequalities

$$w_{\hat{n}} - w_{\hat{n}-1} \leqq K \leqq w_{\hat{n}+1} - w_{\hat{n}} .$$

Assume now that the rate q is unknown to the controller, and write $\hat{\sigma} = \hat{\sigma}(q)$. Let N_T denote the number of arrivals into the system up to time T. Then the ratio

$$q_T^* = N_T/T$$

is the maximum likelihood estimate of q. A natural way of defining an adaptive control is to set

$$Z_t = (\hat{\sigma}(q_t^*))_{X_t} , \quad t \geqq 0 .$$

Since

$$\lim_{T\to\infty} q_T^* = q \quad \text{a.s.},$$

(20) holds with $\hat{\sigma} = \hat{\sigma}(q)$. It can be shown that the other hypotheses of Theorem 5 are fulfilled as well.

(Received March 25, 1986.)

REFERENCES

[1] B. M. Brown: Martingale Central Limit Theorems. Ann. Math. Statist. *42* (1971), 59—66.

[2] A. Hordijk: Dynamic Programming and Markov Potential Theory. Math. Centrum, Amsterdam 1974.

[3] P. Mandl: On the adaptive control of countable Markov chains. Prob. Theory, Banach Center Publications, Vol. 5, 159—173, Warsaw 1979.
[4] P. Mandl: Martingale Methods in Discrete State Random Processes. Supplement to the Journal Kybernetika *18* (1982).
[5] Mª R. Romera: Adaptive Control of Markov Processes with Countable State Space. Doctoral Thesis (in Spanish). Universidad Complutense, Madrid 1985.

RNDr. Petr Mandl, DrSc., matematicko-fyzikální fakulta Univerzity Karlovy (Faculty of Mathematics and Physics, Charles University), Sokolovská 83, 186 00 Praha 8. Czechoslovakia.

Prof. Dr. Mª Rosario Romera-Ayllón, Faculty of Informatics, Polytechnical University of Madrid, Km. 7 Carretera de Valencia, 28031 Madrid. Spain.