KYBERNETIKA — VOLUME 7 (1971), NUMBER 1

On the Variance in Controlled Markov Chains

PETR MANDL

The variance of the reward in an absorbing Markov chain and the asymptotic variance of the reward in an ergodic Markov chain are investigated. Attention is payed to optimal homogeneous Markovian controls which minimize the variance.

1. NOTATIONS

Notations introduced in [4] are followed in the present paper and references to [4] are made on some places. However, both the notations and the references are accompanied by comments which make the use of [4] avoidable.

Let a controlled Markov chain with state space $I = \{1, 2, \ldots, r\}$ be given by two systems of matrices

$$\|p(j, k; z)\|_{j,k=1}^{r}, \quad \|c(j, k; z)\|_{j,k=1}^{r}, \quad z \in J = \{1, 2, \ldots, s\}.$$

$p(j, k; z)$ is the transition probability from state j into state k under control parameter value z. $c(j, k; z)$ is the reward from such transition. A control of the chain is identified with a sequence of functions

$$\omega = \{z_n(j_0, \ldots, j_n),\ n = 0, 1, \ldots\}.$$

$z_n(j_0, \ldots, j_n)$ is the control parameter value which is chosen at time n following the occurrence of states $j_0, \ldots, j_n$. ω is called a homogeneous Markovian control if $z_n(j_0, \ldots, j_n) = z(j_n)$, $n = 0, 1, \ldots$ We write then $\omega \sim z(j)$. For each initial state j, a control ω together with the transition probabilities define the probability distribution P_j^{ω} of a sequence $\{X_n, n = 0, 1, \ldots\}$ of random variables describing the development of the chain under ω. The mathematical expectation with respect to this probability distribution will be denoted by E_j^{ω}.

The reward up to time n is given by

$$C_n = \sum_{m=1}^{n} c(X_{m-1}, X_m; z_{m-1}(X_0, \ldots, X_{m-1})).$$

2 We do not mark the dependence of C_n on ω explicitly. For shortening we shall also omit the arguments $X_0, \ldots, X_k$ and $j_0, \ldots, j_k$ in $z_k(X_0, \ldots, X_k)$ and $z_k(j_0, \ldots, j_k)$, respectively.

2. REWARD UP TO THE FIRST ENTRANCE INTO A CHOSEN SET OF STATES

Let $I_0 \subset I$. Denote by N the first entrance time of the trajectory $\{X_n, n = 0, 1, \ldots\}$ into I_0, i.e.,

$$N = \inf \{n : X_n \in I_0\} .$$

We assume that

$$E_j^\omega N < \infty , \quad j \in I , \tag{1}$$

holds for all homogeneous Markovian controls ω. By Theorem 1 § 8 [4], relation (1) is valid for arbitrary ω. We introduce

$$C = C_N = \sum_{n=1}^{N} c(X_{n-1}, X_n; z_{n-1}) ,$$

the reward up to the first entrance into I_0. Our investigation of the variance of C will be based on a formula for its characteristic function which is presented as Theorem 1.

Note that for an arbitrary homogeneous Markovian control $\tilde{\omega} \sim \tilde{z}(j)$ we have

$$\begin{aligned} E_j^{\tilde{\omega}} e^{i\vartheta C} &= \sum_k p(j, k; \tilde{z}(j))\, e^{i\vartheta c(j,k;\tilde{z}(j))} E_k^{\tilde{\omega}} e^{i\vartheta C} , \quad j \in I - I_0 , \\ E_j^{\tilde{\omega}} e^{i\vartheta C} &= 1 , \quad j \in I_0 , \quad -\infty < \vartheta < \infty . \end{aligned} \tag{2}$$

Moreover, (2) as system of equations for $(E_1^{\tilde{\omega}} e^{i\vartheta C}, \ldots, E_r^{\tilde{\omega}} e^{i\vartheta C})$ has a unique solution for ϑ from a neighbourhood of 0. This follows from the fact that its determinant is non-zero for $\vartheta = 0$ in virtue of (1). Since the solution of (2) is infinitely differentiable at $\vartheta = 0$, we conclude that

$$E_j^{\tilde{\omega}} |C|^k < \infty , \quad k = 1, 2, \ldots, j \in I .$$

For the first two moments we get from (2) the equations

$$\begin{aligned} E_j^{\tilde{\omega}} C &= \sum_k p(j, k; \tilde{z}(j)) \left(c(j, k; \tilde{z}(j)) + E_k^{\tilde{\omega}} C\right) , \quad j \in I - I_0 , \\ E_j^{\tilde{\omega}} C^2 &= \sum_k p(j, k; \tilde{z}(j)) \left(c(j, k; \tilde{z}(j)) + 2c(j, k; \tilde{z}(j))\, E_k^{\tilde{\omega}} C + E_k^{\tilde{\omega}} C^2\right) , \\ & \qquad j \in I - I_0 , \\ E_j^{\tilde{\omega}} C &= E_j^{\tilde{\omega}} C^2 = 0 , \quad j \in I_0 . \end{aligned} \tag{3}$$

Theorem 1. *Let $\tilde{\omega} \sim \tilde{z}(j)$ be a homogeneous Markovian control, ω an arbitrary control. Then*

$$(4)\quad E_j^{\omega} e^{i\vartheta C} = E_j^{\tilde{\omega}} e^{i\vartheta C} + E_j^{\omega} \sum_{n=0}^{N-1} \Psi(X_n, z_n, \vartheta)\, e^{i\vartheta C_n}, \quad j \in I - I_0, \quad -\infty < \vartheta < \infty,$$

where

$$\Psi(j, z, \vartheta) = \sum_k p(j, k; z)\, e^{i\vartheta c(j,k;z)} E_k^{\tilde{\omega}} e^{i\vartheta C} - E_j^{\tilde{\omega}} e^{i\vartheta C}.$$

Proof. Let M be a positive integer. Set $N' = \min\{M, N\}$. To demonstrate the Theorem we shall establish the relation

$$(5)\quad E_j^{\omega} e^{i\vartheta C_{N'}} = E_j^{\tilde{\omega}} e^{i\vartheta C} + E_j^{\omega} \sum_{n=0}^{N'-1} \Psi(X_n, z_n, \vartheta)\, e^{i\vartheta C_n} + E_j^{\omega} \chi_{\{N>M\}} e^{i\vartheta C_M} \left(1 - E_{X_M}^{\tilde{\omega}} e^{i\vartheta C}\right), \quad j \in I - I_0.$$

Here $\chi_{\{N>M\}}$ is the indicator of the random event $\{N > M\}$. (4) follows from (5) by letting $M \to \infty$.

(5) is a special case $(m = 0)$ of

$$(6)\quad E_j^{\omega}\{e^{i\vartheta(C_{N'} - C_m)} \mid X_1 = j_1, \ldots, X_m = j_m\} = E_{j_m}^{\tilde{\omega}} e^{i\vartheta C} + E_j^{\omega}\Big\{ \sum_{n=m}^{N'-1} \Psi(X_n, z_n, \vartheta)\, e^{i\vartheta(C_n - C_m)} + \chi_{\{N>M\}} e^{i\vartheta(C_M - C_m)} \left(1 - E_{X_M}^{\tilde{\omega}} e^{i\vartheta C}\right) \Bigm| X_1 = j_1, \ldots, X_m = j_m\Big\}, \quad j \notin I_0, \ldots, j_m \notin I_0,$$

which will now be verified by induction for $m = M, M - 1, \ldots, 0$. For $m = M$, (6) is obvious. Thus, let (6) hold for some m, $M \geqq m > 0$. Then, substituting from (6) into

$$E_j^{\omega}\{e^{i\vartheta(C_{N'} - C_{m-1})} \mid X_1 = j_1, \ldots, X_{m-1} = j_{m-1}\} = \sum_k p(j_{m-1}, k; z_{m-1})\, e^{i\vartheta c(j_{m-1},k;z_{m-1})} E_j^{\omega}\{e^{i\vartheta(C_{N'} - C_m)} \mid X_1 = j_1, \ldots, X_{m-1} = j_{m-1}, X_m = k\}$$

we obtain

$$E_j^{\omega}\{e^{i\vartheta(C_{N'} - C_{m-1})} \mid X_1 = j_1, \ldots, X_{m-1} = j_{m-1}\} = E_{j_{m-1}}^{\tilde{\omega}} e^{i\vartheta C} - E_{j_{m-1}}^{\tilde{\omega}} e^{i\vartheta C} + \sum_k p(j_{m-1}, k; z_{m-1})\, e^{i\vartheta c(j_{m-1},k;z_{m-1})} E_k^{\tilde{\omega}} e^{i\vartheta C} + E_j^{\omega}\Big\{ \sum_{n=m}^{N'-1} \Psi(X_n, z_n, \vartheta)\, e^{i\vartheta(C_n - C_{m-1})} + \chi_{\{N>M\}} e^{i\vartheta(C_M - C_{m-1})} \left(1 - E_{X_M}^{\tilde{\omega}} e^{i\vartheta C}\right) \Bigm| X_1 = j_1, \ldots, X_{m-1} = j_{m-1}\Big\} =$$

$$= E_{j_{m-1}}^{\tilde{\omega}} e^{i\vartheta C} + E_j^{\omega}\Big\{ \sum_{n=m-1}^{N'-1} \Psi(X_n, z_n, \vartheta)\, e^{i\vartheta(C_n - C_{m-1})} + \chi_{\{N>M\}} e^{i\vartheta(C_M - C_{m-1})} \left(1 - E_{X_M}^{\tilde{\omega}} e^{i\vartheta C}\right) \Bigm| X_1 = j_1, \ldots, X_{m-1} = j_{m-1}\Big\}.$$

By this, the inductive proof of (6) is accomplished. □

Relations for the moments of C are obtained by differentiating (4) and by setting $\vartheta = 0$. Thus, let

$$\Psi_m(j, z) = \frac{\mathrm{d}^m}{\mathrm{d}\vartheta^m} \Psi(j, z, \vartheta)\,|_{\vartheta=0} = \sum_k p(j, k; z)\, E_k^{\tilde{\omega}}(c(j, k; z) + C)^m - E_j^{\tilde{\omega}} C^m .$$

For the first moment $E_j^{\omega} C = u(j; \omega)$ we obtain from (4)

$$u(j; \omega) = u(j; \tilde{\omega}) + E_j^{\omega} \sum_{n=0}^{N-1} \Psi_1(X_n, z_n) . \tag{7}$$

Let $\tilde{\omega} \sim \tilde{z}(j)$ be an optimal control, i.e.

$$u(j; \tilde{\omega}) = \hat{u}(j) = \max_{\omega} u(j; \omega) , \quad j \in I .$$

Then

$$\Psi_1(j; z) = \sum_k p(j, k; z)\,(c(j, k; z) + \hat{u}(k)) - \hat{u}(j) .$$

Moreover, we have the Bellman equation (3 § 8 [4])

$$\hat{u}(j) = \max_z \sum_k p(j, k; z)\,(c(j, k; z) + \hat{u}(k)) , \quad j \in I - I_0 .$$

Hence,

$$0 = \max_z \Psi_1(j, z) , \quad j \in I - I_0 . \tag{8}$$

For $j \in I - I_0$, the set of all z for which $\Psi_1(j, z) = 0$ will be denoted by $J(j)$. (7) and (8) imply the following Corollary, which is a rather immediate consequence of Bellman's principle.

Corollary 1. *ω is an optimal control if and only if*

$$P_j^{\omega}(z_n(X_0, \ldots, X_n) \in J(X_n) , \quad n = 0, 1, \ldots, N-1) = 1 , \quad j \in I - I_0 . \tag{9}$$

We shall now characterize those among controls satisfying (9) under which the variance of C, $E_j^{\omega} C^2 - \hat{u}(j)^2$, is minimal. Thus, let (9) hold for ω. $\tilde{\omega}$ in (4) is assumed to be an optimal control. Then

$$\Psi(j, z, \vartheta) = \tfrac{1}{2}\Psi_2(j, z)\,\vartheta^2 + O(\vartheta^3) \quad \text{for} \quad \vartheta \to 0 , \quad j \in I - I_0 , \quad z \in J(j) . \tag{10}$$

From

$$E_j^{\omega} C^2 = \lim_{\vartheta \to 0} \vartheta^{-2} E_j^{\omega}(\mathrm{e}^{\mathrm{i}\vartheta C} - 2 + \mathrm{e}^{-\mathrm{i}\vartheta C}) =$$

$$= \lim_{\vartheta \to 0} \vartheta^{-2} \{E_j^{\tilde{\omega}}(\mathrm{e}^{\mathrm{i}\vartheta C} - 2 + \mathrm{e}^{-\mathrm{i}\vartheta C}) + E_j^{\omega} \sum_{n=0}^{N-1} (\Psi(X_n, z_n, \vartheta)\,\mathrm{e}^{\mathrm{i}\vartheta C_n} +$$

$$+ \Psi(X_n, z_n, -\vartheta)\,\mathrm{e}^{-\mathrm{i}\vartheta C_n})\} , \quad j \notin I_0 ,$$

we get using (10)

$$(11) \qquad E_j^{\omega} C^2 = E_j^{\hat{\omega}} C^2 + E_j^{\omega} \sum_{n=0}^{N-1} \Psi_2(X_n, z_n), \quad j \notin I_0 .$$

For $j \in I_0$ the validity of (11) is obvious.

(11) implies that the minimization of $E_j^{\omega} C^2 - \hat{u}(j)^2$ amounts to the minimization of $E_j^{\omega} \sum_{n=0}^{N-1} \Psi_2(X_n, z_n)$. We can therefore follow the reasoning which lead to Corollary 1. Introduce

$$\hat{S}(j) = \min \{E_j^{\omega} C^2 : \omega \text{ satisfying } (9)\} .$$

Then

$$\hat{S}(j) - E_j^{\hat{\omega}} C^2 = \min \{E_j^{\omega} \sum_{n=0}^{N-1} \Psi_2(X_n, z_n) : \omega \text{ satisfying } (9)\}$$

fulfils the Bellman equation

$$(12) \quad \hat{S}(j) - E_j^{\hat{\omega}} C^2 = \min_{z \in J(j)} \{\Psi_2(j, z) + \sum_k p(j, k; z) (\hat{S}(k) - E_k^{\hat{\omega}} C^2)\}, \quad j \in I - I_0 .$$

(12) gives after simple transformations

$$(13) \quad \hat{S}(j) = \min_{z \in J(j)} \sum_k p(j, k; z) [c(j, k; z)^2 + 2\hat{u}(k) c(j, k; z) + \hat{S}(k)], \quad j \in I - I_0 ,$$

$$\hat{S}(j) = 0, \quad j \in I_0 .$$

(13) can also be obtained more directly from (3).

Let $J'(j)$ be the set of all $z \in J(j)$ which minimize the expresion on the right-hand side of (13). Using Corollary 1 we get the following Theorem.

Theorem 2. *For $j \in I - I_0$ we have*

$$E_j^{\hat{\omega}} C = \hat{u}(j), \quad E_{j_i}^{\hat{\omega}}(C - \hat{u}(j))^2 = \min \{E_j^{\omega}(C - \hat{u}(j))^2 : E_j^{\omega} C = \hat{u}(j)\}$$

if and only if

$$P_{j_i}^{\hat{\omega}}(z_n(X_0, \ldots, X_n) \in J'(X_n), \quad n = 0, 1, \ldots, N - 1) = 1 .$$

Collorary 2. *There exist homogeneous Markovian controls which minimize the variance of C.*

Remark 1. Relation, analogous to (4), for the discounted reward

$$D = \sum_{m=1}^{\infty} \beta^{m-1} c(X_{m-1}, X_m; z_{m-1}), \quad D_n = \sum_{m=1}^{n} \beta^{m-1} c(X_{m-1}, X_m; z_{m-1}),$$

$$0 < \beta < 1$$

 reads

$$E_j^{\omega} e^{i\vartheta D} = E_j^{\tilde{\omega}} e^{i\vartheta D} + E_j^{\omega} \sum_{n=0}^{\infty} \Psi^{\beta}(X_n, z_n, \beta^n \vartheta)\, e^{i\vartheta D_n}, \quad j \in I, \quad -\infty < \vartheta < \infty .$$

Here

$$\Psi^{\beta}(j, z; \vartheta) = \sum_k p(j, k; z)\, e^{i\vartheta c(j,k;z)} E_k^{\tilde{\omega}} e^{i\vartheta\beta D} - E_j^{\tilde{\omega}} e^{i\vartheta D} .$$

3. MEAN REWARD PER UNIT TIME

In this Section we make the following assumption: For arbitrary homogeneous Markovian control $\omega \sim z(j)$ the states which are recurrent with respect to $\|p(j, k; z(j))\|_{j,k=1}^{r}$ form only one class. A slightly stronger assumption was made in § 9 [4], but it is not difficult to verify that the results from § 9 [4] which will be employed here are valid under the present hypothesis.

Let $\tilde{\omega} \sim \tilde{z}(j)$ be a homogeneous Markovian control. Denote for the sake of brevity

$$\|p(j, k; \tilde{z}(j))\|_{j,k=1}^{r} = \|p_{jk}\|_{j,k=1}^{r} = P, \quad P^n = \|p_{jk}^{(n)}\|_{j,k=1}^{r},$$

$$\|c(j, k; \tilde{z}(j))\|_{j,k=1}^{r} = \|c_{jk}\|_{j,k=1}^{r} .$$

In virtue of the above assumption we have

$$\lim_{n\to\infty} p_{jk}^{(n)} = \pi_k, \quad k \in I, \tag{14}$$

and hence, the limites

$$\Theta(\tilde{\omega}) = \lim_{n\to\infty} n^{-1} E_j^{\tilde{\omega}} C_n = \sum_k \sum_l \pi_k p_{kl} c_{kl}$$

exist and are independent of j. Moreover, the limites

$$w_j = \lim_{n\to\infty} (E_j^{\tilde{\omega}} C_n - n\tilde{\Theta}) = \sum_{n=0}^{\infty} \sum_k \sum_l (p_{jk}^{(n)} - \pi_k)\, p_{kl} c_{kl} \tag{15}$$

are finite. Here $\tilde{\Theta} = \Theta(\tilde{\omega})$. In fact, since the convergence in (14) is exponential,

$$\sum_{n=0}^{\infty} |w_j - E_j^{\tilde{\omega}} C_n - n\tilde{\Theta}| < \infty . \tag{16}$$

Let us recall the following proposition (Theorem 3, § 9 [4]): $\tilde{\Theta}$ is the unique number to which there exist $w_1, \ldots, w_r$ such that

$$w_j + \tilde{\Theta} = \sum_k p_{jk}(c_{jk} + w_k), \quad j \in I . \tag{17}$$

The numbers $w_1, \ldots, w_r$ are determined by (17) up to an additive constant. Consequently, (17) together with

(18) $$\sum_j \pi_j w_j = 0$$

determine $w_1, \ldots, w_r$ satisfying (15).

Consider now the variance of C_n. We have

$$E_j^{\tilde{\omega}}(C_n - n\tilde{\Theta})^2 = \sum_k p_{jk} E_k^{\tilde{\omega}}(c_{jk} + C_{n-1} - n\tilde{\Theta})^2 = \sum_k p_{jk}[(c_{jk} - \tilde{\Theta})^2 +$$

$$+ 2(c_{jk} - \tilde{\Theta})\, w_k + E_k^{\tilde{\omega}}(C_{n-1} - \overline{n-1}\tilde{\Theta})^2 + 2(c_{jk} - \tilde{\Theta})\,(E_k^{\tilde{\omega}} C_{n-1} - \overline{n-1}\tilde{\Theta} - w_k)] .$$

Setting

(19) $$\sum_k p_{jk}[(c_{jk} - \tilde{\Theta})^2 + 2(c_{jk} - \tilde{\Theta})\, w_k] = \gamma_j$$

$$2\sum_k p_{jk}(c_{jk} - \tilde{\Theta})\,(E_k^{\tilde{\omega}} C_n - n\tilde{\Theta} - w_k) = \varrho_j^n$$

we can write

$$E_j^{\tilde{\omega}}(C_n - n\tilde{\Theta})^2 = \gamma_j + \varrho_j^{n-1} + \sum_k p_{jk} E_k^{\tilde{\omega}}(C_{n-1} - \overline{n-1}\tilde{\Theta})^2 , \quad n = 1, 2, \ldots$$

From here,

$$E_j^{\tilde{\omega}}(C_n - n\Theta)^2 = \sum_{m=0}^{n-1} \sum_k p_{jk}^{(m)}(\gamma_k + \varrho_k^{n-m-1}) , \quad n = 1, 2, \ldots$$

Note that by (16)

$$\sum_{m=0}^{\infty} |\varrho_k^m| < \infty , \quad k \in I ,$$

and denote

$$\sigma(\tilde{\omega})^2 = \tilde{\sigma}^2 = \sum_k \pi_k \gamma_k .$$

Then

$$E_j^{\tilde{\omega}}(C_n - n\tilde{\Theta})^2 - n\tilde{\sigma}^2 = \sum_{m=0}^{n-1} \sum_k [(p_{jk}^{(m)} - \pi_k)\, \gamma_k + \pi_k \varrho_k^m] +$$

$$+ \sum_{m=0}^{n-1} \sum_k (p_{jk}^{(n-m-1)} - \pi_k)\, \varrho_k^m .$$

Thus again, the limites

(20) $$w_{2j} = \lim_{n\to\infty} [E_j^{\tilde{\omega}}(C_n - n\tilde{\Theta})^2 - n\tilde{\sigma}^2] = \sum_{m=0}^{\infty} \sum_k [(p_{jk}^{(m)} - \pi_k)\, \gamma_k + \pi_k \varrho_k^m]$$

are finite. A characterization of $\tilde{\sigma}^2$ analogous to (17) can be given. Namely, $\tilde{\sigma}^2$ is the unique number such that

(21) $$w_{2j} + \tilde{\sigma}^2 = \gamma_j + \sum_k p_{jk} w_{2k} , \quad j \in I ,$$

for appropriate $w_{21}, \ldots, w_{2r}$. Any such numbers differ from those defined by (20) at most by an additive constant.

The following Theorem contains formulas which are basic for the investigations of the present Section.

Theorem 3. *Let $\tilde{\omega} \sim \tilde{z}(j)$ be a homogeneous Markovian control, ω an arbitrary control. Then*

$$E_j^\omega C_M = M\tilde{\Theta} + w_j - E_j^\omega w_{X_M} + \sum_{n=0}^{M-1} E_j^\omega \varphi_1(X_n, z_n), \tag{22}$$

$$E_j^\omega(C_M - M\tilde{\Theta})^2 = M\tilde{\sigma}^2 + w_{2j} - E_j^\omega w_{2X_M} + 2\sum_{n=0}^{M-1} E_j^\omega \varphi_1(X_n, z_n)(C_n - n\tilde{\Theta}) - \tag{23}$$

$$- 2E_j^\omega(C_M - M\tilde{\Theta})\, w_{X_M} + \sum_{n=0}^{M-1} E_j^\omega \varphi_2(X_n, z_n), \quad j \in I, \quad M = 1, 2, \ldots,$$

where

$$\varphi_1(j, z) = \sum_k p(j, k; z)\left[c(j, k; z) + w_k\right] - \tilde{\Theta} - w_j, \tag{24}$$

$$\varphi_2(j, z) = \sum_k p(j, k; z)\left[(c(j, k; z) - \tilde{\Theta})^2 + 2(c(j, k; z) - \tilde{\Theta})\, w_k + w_{2k}\right] - \tag{25}$$

$$- w_{2j} - \tilde{\sigma}^2.$$

Proof. (22) and (23) are established using induction with respect to M. We shall verify (23). The proof of (22) is similar. Thus, let (23) hold for some M. Note that

$$E_j^\omega \varphi_1(X_M, z_M)(C_M - M\tilde{\Theta}) = \tag{26}$$

$$= E_j^\omega(c(X_M, X_{M+1}; z_M) - \tilde{\Theta} + w_{X_{M+1}} - w_{X_M})(C_M - M\tilde{\Theta}),$$

$$E_j^\omega \varphi_2(X_M, z_M) = E_j^\omega[(c(X_M, X_{M+1}; z_M) - \tilde{\Theta})^2 + \tag{27}$$

$$+ 2(c(X_M, X_{M+1}; z_M) - \tilde{\Theta})\, w_{X_{M+1}} + w_{2X_{M+1}} - w_{2X_M}] - \tilde{\sigma}^2.$$

Hence,

$$E_j^\omega(C_{M+1} - (M + 1)\tilde{\Theta})^2 = E_j^\omega(C_M - M\tilde{\Theta})^2 + \tag{28}$$

$$+ 2E_j^\omega(c(X_M, X_{M+1}; z_M) - \tilde{\Theta})(C_M - M\tilde{\Theta}) + E_j^\omega(c(X_M, X_{M+1}; z_M) - \tilde{\Theta})^2 +$$

$$+ E_j^\omega\varphi_2(x_M, z_M) - 2E_j^\omega(c(X_M, X_{M+1}; z_M) - \tilde{\Theta})\, w_{X_{M+1}} -$$

$$- E_j^\omega w_{2X_{M+1}} + E_j^\omega w_{2X_M} + \tilde{\sigma}^2.$$

Inserting for $E_j^\omega(C_M - M\tilde{\Theta})^2$ from (23) into (28) we get

$$E_j^\omega(C_{M+1} - (M + 1)\tilde{\Theta})^2 = (M + 1)\tilde{\sigma}^2 + w_{2j} - E_j^\omega w_{2X_{M+1}} +$$

$$+ 2\sum_{n=0}^{M} E_j^\omega\, \varphi_1(X_n, z_n)\,(C_n - n\tilde{\Theta}) - 2E_j^\omega(C_{M+1} - (M+1)\,\tilde{\Theta})\, w_{X_{M+1}} +$$

$$+ \sum_{n=0}^{M} E_j^\omega\, \varphi_2(X_n, z_n)\,.$$

Since (23) holds obviously for $M = 0$, the inductive proof is accomplished. □

Note that no special properties of $\tilde{\Theta}$, $w_1, \ldots, w_r$, $\tilde{\sigma}^2$, $w_{21}, \ldots, w_{2r}$ were used in the proof of Theorem 3.

Let

(29) $$\tilde{\Theta} = \max\{\Theta(\varpi) : \varpi \text{ homogeneous Markovian}\}\,.$$

Then for appropriate $\tilde{\omega} \sim \tilde{z}(j)$ holds $\tilde{\Theta} = \Theta(\tilde{\omega})$ together with the Bellman equation (Theorem 4, § 9 [4])

(30) $$w_j + \tilde{\Theta} = \max_z \sum_k p(j, k; z)\,(c(j, k; z) + w_k)\,, \quad j \in I\,,$$

or

(31) $$\max_z \varphi_1(j, z) = 0\,, \quad j \in I\,.$$

Denote by $K(j)$ the set of all z for which $\varphi_1(j, z) = 0$.

Theorem 4. *Let* (29) *hold and let* ω *be an arbitrary control. Then* $E_j^\omega C_M - M\tilde{\Theta}$, $M = 1, 2, \ldots$, *is bounded from above. It is bounded if and only if*

(32) $$\sum_{n=0}^{\infty} P_j^\omega(z_n(X_0, \ldots, X_n) \notin K(X_n)) < \infty\,.$$

Proof. We shall employ (22). From (31),

$$E_j^\omega C_M - M\tilde{\Theta} \leqq w_j - E_j^\omega w_{X_M} \leqq w_j - \min_k w_k\,.$$

Moreover, $E_j^\omega C_M - M\tilde{\Theta}$, $M = 1, 2, \ldots$, is bounded if and only if

(33) $$\sum_{n=0}^{\infty} E_j^\omega\, \varphi_1(X_n, z_n) > -\infty\,.$$

(33) is equivalent to (32). □

The conjecture that all controls satisfying $P_j^\omega(z_n \in K(X_n),\, n = 0, 1, 2, \ldots) = 1$ maximize the mean reward per unit time was presented to the author by Z. Koutský.

Corollary 3. *If* (32) *holds, then*

$$\max_{\varpi} E_j^{\varpi} C_M - E_j^\omega C_M\,, \quad M = 1, 2, \ldots,$$

is bounded.

Hence, under ω, the expected reward up to time M is sufficiently close to the maximal expected reward for all M.

Let us now investigate the asymptotic behaviour of $E_j^\omega(C_M - M\tilde{\Theta})^2$ for controls satisfying (32). The investigation will be based on (23). Analogously to (15), (17), for arbitrary $\hat{\omega} \sim \hat{z}(j)$ holds

$$\lim_{M\to\infty} \sum_{n=0}^{M-1} E_j^{\hat{\omega}} \varphi_2(X_n, z_n) - M\,\eta(\hat{\omega}) = v_j\,, \quad j \in I\,,$$

where

$$v_j + \eta(\hat{\omega}) = \varphi_2(j, \hat{z}(j)) + \sum_k p(j, k; \hat{z}(j))\, v_k\,, \quad j \in I\,.$$

Set

$$\eta(\hat{\omega}) = \hat{\eta}\,, \quad \psi_1(j, z) = \varphi_2(j, z) + \sum_k p(j, k; z)\, v_k - v_j - \hat{\eta}\,, \quad j \in I,\ z \in J\,.$$

Then by (22)

$$\text{(34)} \qquad \sum_{n=0}^{M-1} E_j^\omega \varphi_2(X_n, z_n) = M\hat{\eta} + v_j - E_j^\omega v_{X_M} + \sum_{n=0}^{M-1} E_j^\omega\, \psi_1(X_n, z_n)\,.$$

The insertion of (34) into (23) yields

$$\text{(35)} \qquad E_j^\omega(C_M - M\tilde{\Theta})^2 = M(\tilde{\sigma}^2 + \hat{\eta}) + w_{2j} + v_j - E_j^\omega(w_{2X_M} + v_{X_M})$$
$$+\, 2\sum_{n=0}^{M-1} E_j^\omega \varphi_1(X_n, z_n)\,(C_n - n\tilde{\Theta}) - 2E_j^\omega(C_M - M\tilde{\Theta})\, w_{X_M} + \sum_{n=0}^{M-1} E_j^\omega \psi_1(X_n, z_n)\,.$$

Let $\hat{\omega} \sim \hat{z}(j)$ be chosen so that

$$\text{(36)} \qquad \hat{z}(j) \in K(j)\,, \quad j \in I\,,$$

$$\text{(37)} \qquad v_j + \hat{\eta} = \min_{z\in K(j)} \{\varphi_2(j, z) + \sum_k p(j, k; z)\, v_k\}\,, \quad j \in I\,.$$

Then $\Theta(\hat{\omega}) = \tilde{\Theta}$ and (see Theorems 3 and 4, § 9 [4]),

$$\hat{\eta} = \min\{\eta(\bar{\omega}) : \bar{\omega} \sim \bar{z}(j)\,,\ \ \bar{z}(j) \in K(j)\,,\ \ j \in I\}\,.$$

(37) can be written in the form

$$\text{(38)} \qquad \min_{z\in K(j)} \psi_1(j, z) = 0\,, \quad j \in I\,.$$

Setting $\omega = \hat{\omega}$ in (35) we get

$$E_j^{\hat{\omega}}(C_M - M\tilde{\Theta})^2 = M(\tilde{\sigma}^2 + \hat{\eta}) + w_{2j} + v_j - E_j^\omega(w_{2X_M} + v_{X_M}) -$$
$$- 2E_j^\omega(C_M - M\tilde{\Theta})\, w_{X_M}\,.$$

From here we infer without difficulties that

$$\hat{\sigma}^2 = \sigma(\hat{\omega})^2 = \tilde{\sigma}^2 + \hat{\eta}\,.$$

Theorem 5. *Let* (29), (37) *hold. Let* ω *be a control satisfying* (32). *Then, for* $M \to \infty$,

$$E_j^{\omega}(C_M - M\tilde{\Theta})^2 = M\hat{\sigma}^2 + \sum_{n=0}^{M-1} E_j^{\omega}\,\psi_1(X_n, z_n) + O(\sqrt{M})\,, \quad j \in I\,. \tag{39}$$

Proof. Consider (35). Note that (32) implies

$$\sum_{n=0}^{\infty} \sqrt{(E_j^{\omega}\,\varphi_1(X_n, z_n)^2)} < \infty\,,$$

and set

$$A = 2\Big(\sum_{n=0}^{\infty} \sqrt{(E_j^{\omega}\,\varphi_1(X_n, z_n)^2)} + \max_k |w_k|\Big)\,,$$

$$B = \max_{k,l}\,(w_{2k} + v_k - w_{2l} - v_l)\,.$$

Employing Schwartz Inequality one gets from (35)

$$E_j^{\omega}(C_M - M\tilde{\Theta})^2 - A\sqrt{[\max_{1 \leq n \leq M} E_j^{\omega}(C_n - n\tilde{\Theta})^2]} - B \leq M\hat{\sigma}^2 + \tag{40}$$

$$+ \sum_{n=0}^{M-1} E_j^{\omega}\psi_1(X_n, z_n) \leq E_j^{\omega}(C_M - M\tilde{\Theta})^2 + A\sqrt{[\max_{1 \leq n \leq M} E_j^{\omega}(C_n - n\tilde{\Theta})^2]} + B\,.$$

(38) and (32) imply

$$\max_{1 \leq n \leq M}\Big[n\hat{\sigma}^2 + \sum_{m=0}^{n-1} E_j^{\omega}\psi_1(X_m, z_m)\Big] \leq M\hat{\sigma}^2 + \sum_{n=0}^{M-1} E_j^{\omega}\psi_1(X_n, z_n) + C\,, \quad M = 1, 2, \ldots,$$

where C is a constant independent of M. Hence, the chain of inequalities as in (40) holds with $E_j^{\omega}(C_M - M\tilde{\Theta})^2$ replaced by $\max_{1 \leq n \leq M} E_j^{\omega}(C_n - n\tilde{\Theta})^2$ and B replaced by $B + C$. From this, using elementary reasoning, one obtains

$$\sqrt{[\max_{1 \leq n \leq M} E_j^{\omega}(C_n - n\tilde{\Theta})^2]} = O(\sqrt{M})\,, \quad M \to \infty\,. \tag{41}$$

(40), (41) imply (39). □

Corollary 4. *For any control* ω *satisfying* (32) *we have*

$$E_j^{\omega}(C_M - M\tilde{\Theta})^2 \leq M\hat{\sigma}^2 + O(\sqrt{M})\,, \quad M \to \infty\,.$$

This follows from (39), (38), (37). On the other hand for the homogeneous Markovian control $\hat{\omega}$

$$E_j^{\hat{\omega}}(C_M - M\tilde{\Theta})^2 = M\hat{\sigma}^2 + O(1)\,, \quad M \to \infty\,,$$

according to (20).

(Received September 3, 1970.)

REFERENCES

[1] R. Bellman: Dynamic programming. Princeton 1957.
[2] R. A. Howard: Dynamic programming and Markov processes. New York, London 1960.
[3] J. G. Kemeny, J. L. Snell: Finite Markov chains. Princeton 1960.
[4] P. Mandl: Controlled Markov chains (in Czech). Kybernetika *6* (1969), Supplement, 1—74.

VÝTAH

O rozptylu v řízených Markovových řetězcích

PETR MANDL

Práce se týká řízených Markovových řetězců s konečným počtem stavů i hodnot parametru řízení. První část je věnována celkovému výnosu do prvého dosažení zvolené množiny stavů. Je odvozen vztah (4) pro charakteristické funkce výnosu. Z něj vyplývá charakterizace optimálních řízení, minimalizujících rozptyl výnosu, obsažená ve větě 2. Výklad v části o průměrném výnosu na jednotku času je založen na rovnostech (22), (23), které jsou nepřímým důsledkem (4). Věta 4 obsahuje nutnou a postačující podmínku pro to, aby při řízení ω rozdíl mezi očekávaným výnosem a maximálním možným výnosem do libovolného časového okamžiku byl ohraničen. Pro taková řízení je odvozeno asymptotické vyjádření rozptylu (39). Z něj plyne zejména, že minimálního asymptotického rozptylu je dosaženo při homogenním markovském řízení.

Dr. Petr Mandl, DrSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Vyšehradská 49, Praha 2.