KYBERNETIKA — VOLUME 10 (1974), NUMBER 2

A Diffusion Approximation in the Ruin Problem for a Controlled Markov Chain

PHAM VAN KIEU

The reward from a controlled Markov chain is approximated by a diffusion process. From a control policy maximizing its expected discounted trajectory under a penalty for reaching zero a control of the original Markov chain is derived.

The ruin problem in controlled Markov process was considered by Z. Koutský in [4]. In the present paper a diffusion appproximation is used to calculate controls taking the ruin probability into account. In Part 1 the problem is defined and the proposed solution is explained. In Part 2 a limit theorem is given which confirms the legitimacy of the approximations employed.

1. THE STATEMENT OF THE PROBLEM

Let $\{X_n, n = 0,1,\ldots\}$ denote the trajectory of a controlled Markov chain with transition probabilities

$$(1) \qquad p(i, j; z), \quad z \in \mathscr{Z}(i) \quad i, j \in I .$$

Here I is the finite state space of the chain, $\mathscr{Z}(i)$ the set of control parameter values in state I, $i \in I$. $\mathscr{Z}(i)$, $i \in I$, are assumed to be closed and bounded in R^s. $p(i, j; z)$ is the transition probability from state i into state j under control parameter value z. Further, let $c(i, j; z)$ denote the reward the controller gets from such transition. The functions

$$c(i, j; z), \quad z \in \mathscr{Z}(i), \quad i, j \in I,$$

as well as the probabilities (1), are assumed to be continuous in z.

Let Z_n be the control parameter the controller selects after n steps. $\{Z_n, n = 0, 1, \ldots\}$ is thus a sequence of random variables depending on the past trajectory. Suppose that the controller posseses an initial capital C_0. His capital after M steps includes the reward from the chain, and equals therefore

$$C_M = C_0 + \sum_{m=0}^{M-1} c(X_m, X_{m+1}; Z_m), \quad M = 1, 2, \ldots$$

Introduce $R = \inf\{M : C_M \leqq 0\}$. If $R < \infty$, we say that the controller was ruined after R steps. To balance the change of being ruined and his aim to maximize the reward when selecting the control policy, the controller employs the criterion

$$E\{C_0 + \sum_{m=0}^{R-1} d^{m+1} c(X_m, X_{m+1}; Z_m) - Nd^R\}, \tag{2}$$

where d is a discount factor, $0 \leqq d < 1$, and $N > 0$ is a penalty for the ruin

From the Markovian property it follows that after n steps, (X_n, C_n) contains sufficient information for controlling the chain in an optimal way according to criterion (2). The controller thus looks for a function

$$z(i, C), \quad i \in I, \quad C \in (0, \infty), \tag{3}$$

such that the expectation (2) is maximal for $Z_n = z(X_n, C_n)$, $n = 0, 1, \ldots, R - 1$.

Let Ω be the set of all functions $\omega \sim z(i)$ mapping $i \in I$ into $\mathscr{Z}(i)$. Ω is the set of stationary controls. (3) can be written as

$$\omega(C), \quad C \in (0, \infty). \tag{4}$$

To obtain a diffusion approximation for C_n, $n = 0, 1, \ldots,$ introduce the duration τ of one step in the chain. Thus, the controller's capital at time t equals

$$\mathscr{C}_t = C_{[t/\tau]} + \{t/\tau\} (C_{[t/\tau]+1} - C_{[t/\tau]}). \tag{5}$$

In (5), $[a]$ $\{a\}$ denote the integral part and the fractional part of a, respectively. Linear interpolation is used to make $\mathscr{C}_t$ continuous. Assume first that (4) defines a stationary control, i.e.

$$\omega(C) \equiv \omega \sim z(i), \quad C \in (0, \infty).$$

Then $\{X_n, n = 0, 1, \ldots\}$ is a homogeneous Markov chain with transition probabilities

$$\|p(i, j; z(i))\|_{i,j \in I}. \tag{6}$$

Thorought the paper we shall make the following hypothesis.

Assumption. For arbitrary $\omega \in \Omega$ the states which are recurrent with respect to the transition probability matrix (6) form only on irreducible set.

From the central limit theorem for Markov chains follows that C_n is asymptotically normally distributed $N(\Theta(\omega)\, n, \sigma^2(\omega)n)$ as $n \to \infty$. Denote $1/\sqrt{\tau} = k$. Let

$$(5') \qquad C_0 \approx k\,, \quad \Theta(\omega) \approx 1/k\,, \quad \sigma^2(\omega) \approx 1\,,$$

where k is fairly large. $\mathscr{C}_{t+\Delta} - \mathscr{C}_t$ will be approximately normal $N(\Theta(\omega)\, k^2\Delta$, $\sigma^2(\omega)\, k^2\Delta)$ for $k^2\Delta$ large. Thus we expect that the evolution of $\mathscr{C}_t/k$ will be sufficiently closely described by the stochastic differential equation

$$(6') \qquad \mathrm{d}\gamma_t = \Theta(\omega)\, \mathrm{k}\, \mathrm{d}t + \sigma(\omega)\, \mathrm{d}W_t\,, \quad t \geqq 0\,,$$

where $\{W_t, t \geqq 0\}$, is a standartized Wiener process.

In Part 2 of the paper we give a limit theorem establishing the convergence of $\mathscr{C}_t/k$ to the solution of (6′) in the non-stationary case provided that $\omega(C)$ is continuous. ω in (6) then depends on $k\gamma_t$. To approximate the criterion, we set $d = \exp(-\lambda/k^2)$ and consider, instead of (2),

$$(7) \qquad E\left(\gamma_0 + \int_0^\zeta \mathrm{e}^{-\lambda t}\, \mathrm{d}\gamma_t - \nu\, \mathrm{e}^{-\lambda\zeta}\right) = E\left(\lambda \int_0^\zeta \mathrm{e}^{-\lambda t}\, (\gamma_t + \nu)\, \mathrm{d}t\right) - \nu\,.$$

where

$$\zeta = \inf\{t : \gamma_t \leqq 0\}\,, \quad \nu = N/k\,.$$

The original problem is thus converted into the problem of controlling the diffusion (6′) in such way that (7) is maximal. The following recipe can be found in the literature ([5], [7]). Solve

$$\max_{\omega\in\Omega}\left\{\tfrac{1}{2}\sigma^2(\omega)\,\frac{\mathrm{d}^2 v}{\mathrm{d}\gamma^2} + \Theta(\omega)\, k\,\frac{\mathrm{d}v}{\mathrm{d}\gamma}\right\} - \lambda v + \lambda(\gamma + \nu) = 0\,,$$

$$v(0) = 0\,, \quad v(\gamma) = O(\gamma) \quad \text{as} \quad \gamma \to \infty\,.$$

Let a control $\hat{\omega}(\gamma)$, $\gamma \in [0, \infty)$, be such that

$$\tfrac{1}{2}\sigma^2(\hat{\omega}(\gamma))\,\frac{\mathrm{d}^2 v(\gamma)}{\mathrm{d}\gamma^2} + \Theta(\hat{\omega}(\gamma))\, k\,\frac{\mathrm{d}v(\gamma)}{\mathrm{d}\gamma} - \lambda v(\gamma) + \lambda(\gamma + \nu) = 0\,.$$

Then $\hat{\omega}(\gamma)$ is optimal.

If the conditions for the validity of the diffusion approximation (i.e. essentially the order relations (5′) and the continuity of $\hat{\omega}(\gamma)$) are fulfilled, then a choice of (4) which is nearly optimal with respect to the criterion (2) is given by

$$\omega(C) = \hat{\omega}(C/k)\,, \quad C \in [0, \infty)\,,$$

or

$$Z_n = \hat{z}(X_n, C_n/k)\,, \quad n = 0, 1 \ldots$$

Further, the maximal value of (2) is approximately $kv(C_0/k) - N$.

Finally let us mention that to determine $\Theta(\omega)$, $\sigma^2(\omega)$ for given $\omega \sim z(i)$ one has to solve

$$(8) \qquad \sum_j p(i,j;z(i))\,[c(i,j;z(i)) + w_j] - w_i - \Theta = 0\,, \quad i \in I\,,$$

and

$$(9) \qquad \sum_j p(i,j;z(i))\,[(c(i,j;z(i)) - \Theta)^2 + 2(c(i,j;z(i)) - \Theta)\,w_j + w_{2j}] - \\ - w_{2i}\,\sigma^2 - = 0\,, \quad i \in I\,,$$

for the unknowns Θ, w_i, $i \in I$, and σ^2, w_{2i}, $i \in I$, respectively [1], [6].

2. THE LIMIT THEOREM

Let the reward functions c depend on an auxiliary parameter $k = 1, 2, \ldots,$ and denote them by

$$c(i,j;z,k)\,, \quad z \in \mathscr{Z}(i)\,, \quad i,j \in I\,, \quad k = 1, 2, \ldots$$

The functions c are assumed to be uniformly bounded. To each k there corresponds a controlled Markov chain with transition probabilities (1), as described in Section 1. To mark the dependence on the parameter, we shall add the index k to the symbols like X_n^k, C_n^k, Z_n^k etc.

Introduce

$$\sum_j p(i,j;z)\,c(i,j;z,k) = r_1(i,z;k)\,, \quad z \in \mathscr{Z}(i)\,, \quad i \in I, k = 1, \ldots,$$

$$\sum_j p(i,j;z)\,c(i,j;z,k)^2 = r_2(i,z;k)\,, \quad z \in \mathscr{Z}(i)\,, \quad i \in I\,, \quad k = 1, 2, \ldots.$$

Assume

$$(10) \qquad \left.\begin{aligned} \lim_{k\to\infty} k r_1(i,z;k) = \varrho_1(i,z) \\ \lim_{k\to\infty} k r_2(i,z;k) = \varrho_2(i,z) \end{aligned}\right\} \text{uniformly in } z \in \mathscr{Z}(i)\,, \quad i \in I\,.$$

Theorem. *Let, for $i \in I$, $z(i,\gamma)$ be a continuous mapping of $\gamma \in [0, \infty)$ into $\mathscr{Z}(i)$. Assume that $\{X_n^k, n = 0, 1, \ldots\}$, $k = 1, 2, \ldots$, is controlled by $Z_n^k = z(X_n^k, C_n^k/k)$, $n = 0, 1 \ldots$. Let $T > O$. Denote by $\mathscr{P}_T^k$ the probability distribution of*

$$\{\gamma_t^k = k^{-1}[C_{[tk^2]}^k + \{tk^2\}\,(C_{[tk^2]+1}^k - C_{[tk^2]}^k)]\,, \quad t \in [0, T]\}$$

in the space γ of continuous functions on $[0, T]$.

If (10) holds, and $\lim_{k\to\infty} C_0^k/k = \bar{\gamma}$, then $\mathscr{P}_T^k$ as $k \to \infty$ converges weakly to the *probability distribution* $\mathscr{P}_T$ *of the Markov process* $\{\gamma_t, t \in [0, T]\}$, *satisfying the stochastic differential equation*

$$\mathrm{d}\gamma_t = \Theta(\gamma_t)\,\mathrm{d}t + \sigma(\gamma_t)\,\mathrm{d}W_t\,, \quad t \geqq 0\,; \quad \gamma_0 = \bar{\gamma}\,. \tag{11}$$

$\{W_t, t \geqq 0\}$ *is a standartized Wiener process.* $\Theta(\gamma)$ *and* $\sigma(\gamma)$ *are obtained from the equations*

$$\varrho_1(i, z(i, \gamma)) + \sum_j p(i, j; z(i, \gamma))\, w_j - w_i - \Theta = 0\,, \quad i \in I\,,$$

$$\varrho_2(i, z(i, \gamma)) + \sum_j p(i, j; z(i, \gamma))\, w_{2j} - w_{2i} - \sigma^2 = 0\,, \quad i \in I\,,$$

for the unknowns Θ, w_i, $i \in I$, σ^2, w_{2i}, $i \in I$.

The proof of the theorem uses the methods developed in [6], the tightness of probability measures ([2]), and Doob's Theorem 3.3 ([3]). The course of the proof will be outlined in the subsequent four paragraphs.

a) Solve

$$r_1(i, z(i, \gamma); k) + \sum_j p(i, j; z(i, \gamma))\, w_j^k(\gamma) - w_i^k(\gamma) - \Theta^k(\gamma) = 0\,, \quad i \in I\,,$$

$$\sum_j p(i, j; z(i, \gamma))\,[(c(i. j; z(i, \gamma), k) - \Theta^k(\gamma))^2 + 2w_j^k(\gamma)\,(c(i, j; z(i, \gamma), k) - \Theta^k(\gamma)) +$$
$$+ w_{2j}^k(\gamma)] - w_{2i}^k(\gamma) - \sigma_k^2(\gamma) = 0\,, \quad i \in I\,.$$

Introduce

$$M_n^k = C_n^k - \sum_{m=0}^{n-1} \Theta^k(C_m^k/k) + \sum_{m=0}^{n-1} [w_{X_{m+1}}^k(C_m^k/k) - w_{X_m}^k(C_m^k/k)] \tag{12}$$

or

$$M_n^k = C_n^k - \sum_{m=0}^{n-1} \Theta(C_m^k/k) + \sum_{m=1}^{n-1} [w_{X_m}^k(C_{m-1}^k/k) - w_{X_m}^k(C_m^k/k)] +$$
$$+ w_{X_n}^k(C_{n-1}^k/k) - w_{X_0}^k(C_0^k/k)\,, \quad n = 1, 2, \ldots$$

Then $\{M_n^k, n = 1, 2, \ldots\}$ is a martingale with respect to $\{\mathscr{F}_n^k, n = 1, 2, \ldots\}$, where $\mathscr{F}_n^k$ denotes the Borel field of random events defined on $\{X_0^k, X_1^k, \ldots, X_n^k\}$.

Furthermore,

$$E^k\{(M_{n+l}^k - M_n^k)^2\,|\mathscr{F}_n^k\} = E^k\{w_{2X_0}^k(C_n^k/k) - w_{2X_{n+l}}^k(C_{n+l-1}^k/k) + \tag{13}$$
$$+ \sum_{m=n+1}^{n+l-1} [w_{2X_m}^k(C_m^k/k) - w_{2X_m}^k(C_{m-1}^k/k)] + \sum_{m=n}^{n+l-1} \sigma_k^2(C_m^k/k)\,|\mathscr{F}_n^k\}\,,$$
$$n = 0, 1 \ldots\,, \quad l = 1, 2, \ldots$$

b) Set

$$\mu_t^k = k^{-1}[M_{[tk^2]}^k + \{tk^2\}\,(M_{[tk^2]+1}^k - M_{[tk^2]}^k)]\,, \quad t \in [0, T]\,.$$

Denote by $\mathscr{Q}_T^k$ the probability distribution of $\{\mu_t^k,\ t \in [0, T]\}$ on γ. The sequence $\{\mathscr{Q}_T^k,\ k = 1, 2, \ldots\}$ is tight. In fact, consider, for given $\varepsilon > 0$, the following limit

$$\lim_{\delta\to 0}\ \lim_{k\to\infty} P^k\{\sup_{|s-t|<\delta} |\mu_s^k - \mu_t^k| > \varepsilon\}\,.$$

It holds

$$P^k\left\{\sup_{|s-t|<\delta} |\mu_s^k - \mu_t^k| > \varepsilon\right\} \leqq \sum_{l<T/\delta} p^k\left\{\sup_{l\delta\leqq s\leqq(l+1)\delta} |\mu_s^k - \mu_{l\delta}^k| > \frac{\varepsilon}{4}\right\},$$

$$P^k\left\{\sup_{l\delta\leqq s\leqq(l+1)\delta} |\mu_s^k - \mu_{l\delta}^k| > \frac{\varepsilon}{4}\right\} \leqq P^k\left\{\max_{a_l\leqq r\leqq a_{l+1}} \Big|\sum_{j=a_l+1}^{r} Y_j^k\Big| > \frac{\varepsilon}{4}\right\} \leqq$$

$$\leqq \left(\frac{4}{k\varepsilon}\right)^4 E\Big(\max_{a\leqq r\leqq a_{l+1}} \Big|\sum_{j=a_l+1}^{r} Y_j^k\Big|\Big)^4,$$

where $a_l = [l\delta k^2]$.

To estimate the last term calculate

$$E\Big(\sum_{j=a_l+1}^{a_{l+1}} Y_j^k\Big)^4 = \sum_{a_l\leqq i,j,m,n\leqq a_{l+1}} E(Y_i^k Y_j^k Y_m^k Y_n^k)\,.$$

If the largest index is not matched by any other, then, by $E(Y_n^k|\mathscr{F}_{n-1}^k) = 0$, the term vanishes; hence

$$E\Big(\sum_{j=a_l+1}^{a_{l+1}} Y_j^k\Big)^4 = \sum_{m=a_l+1}^{a_{l+1}} E(Y_m^k)^4 + 4\sum_{a_l\leqq i<m\leqq a_{l+1}} E\{Y_i^k (Y_m^k)^3\} +$$

$$+ 6\sum_{m=2}^{[\delta k_2]} E\Big\{\Big(\sum_{j=a_l+1}^{a_l+m-1} Y_j^k\Big)^2 (Y_{a_l+m}^k)^2\Big\} \leqq \delta^2 k^4\,\text{const}\,.$$

From the martingale inequality (Doob's Theorem 3.4, p. 317).

$$E\{\max_{a_l\leqq r\leqq a_{l+1}} |M_r^k - M_{a_l}^k|^\nu\} \leqq \left(\frac{\nu}{\nu-1}\right)^4 E\{|M_{a_{l+1}}^k - M_{a_l}^k|^\nu\}\,,$$

we get

$$E\{\max_{a_l\leqq r\leqq a_{l+1}} \Big(\sum_{j=a_l+1}^{r} Y_j^k\Big)^4\} \leqq (4/3)^4 k^4 \delta^2\,.\,\text{const}\,.$$

Consequently,

$$P^k\{\sup_{|s-t\leqq\delta} |\mu_s^k - \mu_t^k| > \varepsilon\} \leqq \left(\frac{4}{k\varepsilon}\right)^4 \left(\frac{4}{3}\right)^4 k^4\delta^2\,.\,\text{const}\,\frac{T}{\delta} = \text{const}.\,\delta\,.$$

We conclude that

$$\lim_{\delta\to 0}\ \lim_{k\to\infty}\ P^k\{\sup_{|s-t|\leqq\delta} |\mu_s^k - \mu_t^k| > \varepsilon\} = 0\,.$$

Furthermore, from (12), (10) follows

(14) $$\mu_t^k = \gamma_t^k - \int_0^t \Theta^k(\gamma_u^k)\,\mathrm{d}u + \eta_t^k\,, \quad t \geqq 0\,,$$

where

(15) $$\lim_{k\to\infty} P^k\{\sup_{0\leqq t\leqq T} |\eta_t^k| > \varepsilon\} = 0 \quad \text{for} \quad \varepsilon > 0\,.$$

Let $\overline{\mathscr{Q}}_T^k$ be the probability distribution of

$$\mu_s^k - \eta_t^k = \gamma_t^k - \int_0^t \Theta^k(\gamma_u^k)\,\mathrm{d}u\,.$$

We have

$$P^k\left\{\sup_{|s-t|<\delta} \left|\gamma_s^k - \gamma_t^k - \int_t^s \right|\Theta^k(\gamma_u^k)\,\mathrm{d}u\right| < \varepsilon\right\} \leqq$$

$$\leqq P^k\left\{\sup_{|s-t|<\delta} |\mu_s^k - \mu_t^k| > \frac{\varepsilon}{2}\right\} + P^k\left\{\sup_{|s-t|<\delta} |\eta_s^k - \eta_t^k| > \frac{\varepsilon}{2}\right\}.$$

By the above result and by (15) the right hand side converges to zero as $k \to \infty$, $\delta \to 0$. Thus, $\{\overline{\mathscr{Q}}_T^k, k = 1, 2, \ldots\}$ is tight. Similarly, from $P^k\{\sup_{|s-t|\leqq\delta} |\int_t^s \Theta(\gamma_u^k)\,\mathrm{d}u| > \varepsilon/_2\} \to$ $\to 0$ as $\delta \to 0$, it follows that $\{\mathscr{P}_T^k, k = 1, 2, \ldots\}$ is tight.

c) From this we imply that there exist a subsequence $\{\mathscr{P}_T^{k_j}, j = 1, 2, \ldots\}$ of $\{\mathscr{P}_T^k,$ $k = 1, 2, \ldots\}$, $\lim_{j\to\infty} k_j = \infty$, possesing the weak limit $\mathscr{P}_T$. Define

$$\mu_t = \gamma_t - \int_0^t \Theta(\gamma_u)\,\mathrm{d}u\,, \quad t \in [0, T]\,.$$

Then $\{\mu_t, t \in [0, T]\}$ is a martingale on $(\gamma, \mathscr{P}_T)$ with respect to $\{\Phi_t, t \in [0, T]\}$, where Φ_t denotes the Borel field of random events defined on $\{\gamma_s, s \in [0, t]\}$. From (13) trough a passage to the limit follows

(17) $$\mathscr{E}_T\{(\mu_{t+h} - \mu_t)^2 \,|\Phi_t\} = \mathscr{E}_T\left\{\int_t^{t+h} \sigma^2(\gamma_s)\,\mathrm{d}s\,|\Phi_t\right\}, \quad 0 \leqq t < t + h \leqq T.$$

d) Using (17) and

$$\mathscr{E}_T\{(\gamma_{t+h} - \gamma_t)\,|\Phi_t\} = \mathscr{E}_T\left\{\int_t^{t+h} \Theta(\gamma_s)\,\mathrm{d}s\,|\Phi_t\right\}, \quad 0 \leqq t < t + h \leqq T,$$

the assumptions of Doob's Theorem 3.3 ([3] p. 287) are verified for $\{\gamma_t, t \in [0, T]\}$ on $(\gamma, \mathscr{P}_T)$. This shows that $\mathscr{P}_T$ is the probability distribution of a Markov process, satisfying (11). Such probability distribution is unique. Consequently, $\{\mathscr{P}_T^k, k =$ $= 1, 2, \ldots\}$ has only one accumulation point. This together with its tightness implies the assertion of the theorem.

The paper was prepared under the guidance of DrSc. Petr Mandl in the Institute of Information Theory and Automation, Czechoslovak Academy of Sciences.

(Received October 10, 1973.)

REFERENCES

[1] R. Bellman: A Markovian decision process. J. of Math. and Mech. *6* (1957), 679—684.

[2] P. Billingsley: Convergence of probability measures. New York 1968.

[3] J. Z. Doob: Stochastic processes. New York—London 1953.

[4] Z. Koutský: Probability of a service-system's ruin in the case of finite initial capital. In: Trans. of Fourth Prague Conf. on Information Th. etc. Prague 1967, 425—434.

[5] P. Mandl: Analytical treatment of one-dimensional Markov processes. Prague—Heidelberg 1968.

[6] P. Mandl: A connection between controlled Markov chain and martingales. Kybernetika *9* (1973), 237—241.

[7] R. Morton: On the optimal control of stationary diffusion processes with natural boundaries and discounted cost. J. of Appl. Prob. *8* (1971), 561—572.

Pham van Kieu; Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 180 76 Praha 8. Czechoslovakia.