ISSN 2078-3744. Вісник Львів. ун-ту. Серія мех.-мат. 2013. Випуск 78. С. 171–180
Visnyk of the Lviv Univ. Series Mech. Math. 2013. Issue 78. P. 171–180

УДК 519.21

ОБРИВНІ КЕРОВАНІ МАРКОВСЬКІ ПРОЦЕСИ ЗІ СКІНЧЕННИМИ АБО ЗЛІЧЕННИМИ МНОЖИНАМИ СТАНІВ І КЕРУВАНЬ

Павло ШПАК, Ярослав ЄЛЕЙКО

Львівський національний університет імені Івана Франка,
вул. Університетська, 1, Львів, 79000
e-mail: prshpak@gmail.com

Запропоновано новий підхід до введення ймовірнісної міри на просторі шляхів, змінено означення шляху в рамках обривного керованого марковського процесу та функцію краху. Доведено коректність означень та існування оптимальних, рівномірно оптимальних, ε-оптимальних і рівномірно ε-оптимальних стратегій в умовах нової ймовірнісної міри. Доведено достатність простих стратегій у скінченній і зліченній моделях.

Ключові слова: обривний керований марковський процес, оптимальна стратегія, ε-оптимальна стратегія, рівномірно ε-оптимальна стратегія.

1. Означення обривного керованого марковського процесу. Нехай $X_t(t = m, \ldots, n)$ та $A_t(t = m + 1, \ldots, n)$ довільні не більше, ніж зліченні множини.

Означення 1. *Траєкторія* $l = x_m a_{m+1} x_{m+1} \ldots a_t x_t$ *називається шляхом, якщо* $t = n$ *або* $x_t = x^*$.

Множину всіх можливих шляхів позначатимемо $L = X \times (A \times X)^{n-m}$.

Означення 2. *Обривним керованим марковським процесом на інтервалі часу* $[m, n]$ *є* $(X, A, j, p, q, r, c, \mu) \equiv Z^*_\mu$, *де:*

1) $X = \bigcup\limits_{t=m}^{n} X_t$ *– простір станів;*
2) $A = \bigcup\limits_{t=m+1}^{n} A_t$ *– простір керувань;*
3) *відображення проектування* $j : A \to X$, *де* $j(A_{t+1}) = X_t \backslash \{x^*\}, x^* \in X_t$;
4) *розподіл ймовірностей* $p(\cdot|a) \equiv \mathbb{P}(x_t = x|a_t = ax_{t-1})$ *на* $\mathrm{X_t}$ *з обривними точками* $\mathbb{P}(x_{t+1} = x^* \mid a_t = a) = \mathbb{P}(x_{t+1} = x_m \mid a_t = a) \equiv p(x^* |a) \geq 0$;
5) *функція винагороди* $q : A \to \mathbb{R}$;
6) *фінальна плата* $r : X_n \to \mathbb{R}$;

7) функція краху c, задана в обривних точках станів

$$c(x) = \begin{cases} -\sum\limits_{i=m+1}^{t} \sup\limits_{a_i \in A_i} q(a_i), & x = x^*, \\ 0, & x \neq x^*, \end{cases}$$

$x \in X_t$, x^* – обривна точка. Функція краху задана так, що гарантує повне банкрутство – повну втрату накопиченого капіталу або й більше;

8) початковий розподіл μ на X_m. Якщо початковий розподіл μ зосереджений в точці x, то будемо писати Z^*_x.

Процес, що задовольняє (1–7) називатимемо моделлю і позначатимемо Z^*. Наша мета – знайти спосіб керування, за якого максимізується математичне сподівання оцінки шляху l

$$I(l) = \begin{cases} \sum\limits_{t=m+1}^{n} q(a_t) + r(x_n), & \forall t\, x_t \neq x^*, \\ \sum\limits_{t=m+1}^{k} q(a_t) + c(x_k), & x_k = x^*. \end{cases} \tag{1}$$

2. Обривні керовані марковські процеси для скінченних моделей. Якщо задані перехідна функція $p(\cdot\,|a)$ і стратегії $\pi(\cdot\,|h)$, то кожному початковому розподілу μ відповідає розподіл ймовірностей P^* на просторі L, який набуває такого вигляду:

$$P*(l) = \begin{cases} \mu(x_m)\pi(a_{m+1}\,|x_m)p(x_{m+1}\,|a_{m+1})...\pi(a_n\,|h_{n-1})p(x_n\,|a_n)\ , & \forall t\ x_t \neq x^* \\ \mu(x_m)\pi(a_{m+1}\,|x_m)p(x_{m+1}\,|a_{m+1})...\pi(a_k\,|h_{k-1})p(x_k^*\,|a_k)\ , & x_k = x. \end{cases} \tag{2}$$

Математичне сподівання будь-якої функції ξ з простору L набуде такого вигляду:

$$E^*(\xi) = \sum_{l \in L} \xi(l) P^*(l). \tag{3}$$

Прикладом такої функції є оцінка (1) шляху l. Її математичне сподівання позначимо через ω

$$\omega = EI(l). \tag{4}$$

Теорема 1. *Нехай $\{\pi_k\}$ – скінченний або зліченний набір обривних стратегій і γ_k – невід'ємні числа, сума яких дорівнює 1. Якщо для будь-якого початкового розподілу μ будемо використовувати обривну стратегію π_k з ймовірністю γ_k, то отримуємо в просторі шляхів L розподіл ймовірностей P^*, який набуває вигляду*

$$P^* = \sum_k \gamma_k P_k^*, \tag{5}$$

де розподіл P_k^* відповідає обривній стратегії π_k.

Тоді існує обривна стратегія π, якій відповідає розподіл ймовірностей P^* з (5).

Зауваження 1. Отже, при довільному змішуванні стратегій (виборі стратегій випадково з довільним розподілом ймовірностей) ми не розширимо свої можливості, а отримаємо ще деяку стратегію, яка є комбінацією даних.

Оскільки обривна стратегія π описується скінченним набором невід'ємних чисел $\pi\,(a\,|h\,)$. Набори, які задають стратегію, утворюють замкнуту обмежену множину Π в скінченновимірному просторі. Отже, Π – компакт. Функція $\omega(\pi)$ неперервна, бо виражається через $\pi\,(a\,|h\,)$ за допомогою операцій множення та додавання. За теоремою Вейерштрасса, неперервна функція на Π досягає свого максимуму. Обривна стратегія, при якій досягається максимум, є оптимальною для процесу Z^*. З іншого боку, при кожному $x \in X_m$ існує обривна стратегія π_x, оптимальна для процесу Z_x^*.

За набором обривних стратегій π_x хочемо побудувати обривну стратегію π, оптимальну для моделі Z^*.

Природно користуватися постійно стратегією π_x, якщо шлях починається в точці x. Формально

$$\overline{\pi}\,(\cdot\,|h\,) = \pi_{x(h)}\,(\cdot\,|h\,)\,, \tag{6}$$

де $x(h)$ – початковий стан історії h. Звісно, що формула (6) визначає деяку стратегію $\overline{\pi}$, яка буде оптимальною, тобто $\omega(x,\overline{\pi}) = \omega(x,\pi_x) = \upsilon(x),\ \forall x \in X_m$.

Теорема 2. *Будь-яка обривна стратегія $\overline{\pi}$, задана формулою (6), для якої*

$$\omega(x,\overline{\pi}) = \upsilon(x), \quad x \in X_m \tag{7}$$

є рівномірно оптимальною, тобто $\forall\mu\ \sup\limits_{\pi}\omega(\mu,\pi) = \omega(\mu,\overline{\pi})$.

Доведення. З формул (2)–(4) випливає, що $\forall\pi$

$$\omega(\mu,\pi) = \sum_{l\in L} I(l)P^*(l) = \sum_{X_m}\mu\,(x)\,\omega\,(x,\mu). \tag{8}$$

Зокрема, $\omega\,(\mu,\overline{\pi}) = \sum\limits_{X_m}\mu(x)\omega(x,\overline{\pi})$.

Але $\omega\,(x,\pi) \le \omega\,(x,\overline{\pi})\,,\ \forall x \in X_m$, а отже, $\omega(\mu,\pi) \le \omega(\mu,\overline{\pi})$. □

Наслідок 1. *Для рівномірно оптимальної обривної стратегії $\overline{\pi}$ і довільного початкового розподілу μ*

$$\omega\,(\mu) = \sum_{X_m}\mu(x)\upsilon(x) = \mu\upsilon. \tag{9}$$

Теорема 3. *Справджується таке рівняння*

$$\omega(x,\pi) = \sum_{A(x)}\pi(a\,|x)(q(a) + \omega'(p_a,\pi_a)), \tag{10}$$

де $p_a = p(\cdot\,|a)$, $\pi_a\,(\cdot\,|h'\,) = \pi(\cdot\,|yah')\,$, $a \in A_{m+1}$, $y = j(a)$, h – історія в моделі Z'^*. Рівняння (10) називається фундаментальним і виражає оцінку ω довільної стратегії π в моделі Z^* через оцінку ω' деяких стратегій в моделі Z'^*.

Доведення. Згідно з формулою (8) отримуємо

$$\omega'(p_a,\pi_a) = \sum_{X_{m+1}} p(y\,|a)\,\omega'(y,\pi_a). \tag{11}$$

Розглянемо простори шляхів L та L' в моделях Z^* та Z'^*. Нехай P^* – розподіл ймовірностей в L, який відповідає початковому стану x і стратегії x, P_a^* – розподіл ймовірностей в L', який відповідає початковому розподілу p_a і стратегії π_a.

Згідно з формулою (2) і (2) $\forall l' \in L'$ отримуємо

$$I(xal') = q(a) + I(l') \tag{12}$$

$$P^*(xal') = \pi(a\,|x) P_a^*\ (l'). \tag{13}$$

На підставі (3) і (4) одержуємо

$$\omega(x, \pi) = \sum_L P^*(l) I(l) \tag{14}$$

$$\omega'(p_a, \pi_a) = \sum_{L'} P_a^*(l') I(l'). \tag{15}$$

Міра $P^*(l)$ не дорівнює нулю тільки для шляхів, які починаються з точки x, тобто для шляхів вигляду xal'. Тому, підставляючи в (14) значення $I(l)$ з (12) та $P^*(l)$ з (13) і враховуючи (15), отримуємо фундаментальне рівняння (10). □

З фундаментального рівняння (10) випливає така оцінка:

$$\omega(x, \pi) \le \sup_{A(x)} [q(a) + \omega'(p_a, \pi_a)] \le \sup_{A(x)} [q(a) + \upsilon'(p_a)] \tag{16}$$

$\forall x \in X_m$ і $\forall \pi$ (υ' – оцінка моделі Z'^*). Позначимо

$$u(a) = q(a) + \upsilon'(p_a), \quad (a \in A_{m+1}). \tag{17}$$

Будемо називати цю величину – оцінкою керування a.

Згідно з (9) і рівністю $\upsilon(x^*) = c(x^*)$ отримуємо $u = U\upsilon'$, де оператор U переводить функції на точках (не обривних) у функції на керуваннях за такою формулою:

$$Uf(a) = q(a) + \sum_y p(y\,|a) f(y) + p(y^*\,|a) f(y^*)\ , \tag{18}$$

де y – необривні точки, y^* – обривна точка.

Введемо ще оператор V, який переводить функції на керуваннях у функції на точках (не фінальних та не обривних) так:

$$Vg(x) = \sup_{a \in A(x)} q(a). \tag{19}$$

Запишемо нерівність (16), використовуючи оператор $V : \omega(x, \pi) \le Vu(x)$. Тепер візьмемо $\sup\limits_{\pi}$ від лівої та правої частини й отримаємо $\upsilon \le Vu$.

Теорема 4. *Нехай $\pi = \gamma\pi'$ – добуток обривних стратегій γ та π'. Якщо π' рівномірно оптимальна для моделі Z'^*, тоді*

$$u = Vu. \tag{20}$$

Доведення. Для добутку стратегій фундаментальне рівняння (10) набуде такого вигляду:

$$\omega(x, \gamma\pi') = \sum_{A(x)} \gamma(a\,|x)\ u(a). \tag{21}$$

Оскільки π' – рівномірно оптимальна (а вона існує згідно з теоремою 2), то $\omega'(p_a, \pi') = \upsilon(p_a)$, і згідно з виглядом функції u рівняння (21) перетворюється в

$$\omega(x, \gamma\pi') = \sum_{A(x)} \gamma(a\,|x)\ u(a). \tag{22}$$

Якщо для кожного x розподіл $\gamma(\cdot\,|x)$ зосереджений на $\overline{A}(x) \subset A(x)$, де функція $u(a)$ $(a \in A(x))$ досягає свого максимуму, то це рівняння набуде такого вигляду:

$$\omega(x, \gamma\pi') = Vu(x),\ (x \in X_m). \tag{23}$$

□

Наслідок 2. *Оцінка υ моделі Z^* виражається через оцінку υ' моделі Z'^* такими формулами:*

$$\begin{aligned} \upsilon &= Vu, \\ u &= U\upsilon', \end{aligned} \tag{24}$$

де оператори U та V задані формулами (18) та (19), відповідно.

Наслідок 3. *Існує селектор ψ багатозначного відображення $A(x) : X_m \to A_{m+1}$*

$$u(\psi(x)) = \upsilon(x). \tag{25}$$

Доведення. За $\gamma(\cdot\,|x)$ можна прийняти розподіл, зосереджений в одній точці $\psi(x) \in \overline{A}(x)$. □

Наслідок 4. *Якщо π' рівномірно оптимальна для моделі Z'^* і селектор ψ – такий, як в Наслідку 3, то обривна стратегія $\psi\pi'$ – рівномірно оптимальна для процесу Z^*.*

Не обмежуючи загальності, можна вважати, що в цій моделі Z^* $m = 0$. Розглянемо моделі $Z_0^*, Z_1^*, ...Z_n^*$, де $Z^* = Z_0^*$ і Z_t^* – похідна модель від Z_{t-1}^*. Оцінки υ та u моделі Z_t^* позначимо υ_t і u_{t+1}, відповідно (υ_t визначена на X_t, u_{t+1} визначена на A_{t+1}). Функцію винагороди q і перехідну функцію p на A_t позначимо q_t і p_t, відповідно.

Згідно з попередніми результатами оцінки υ_t і u_t пов'язані між собою такими співвідношеннями:

$$\begin{aligned} \upsilon_{t-1} &= Vu_t \\ u_t &= U\upsilon_t \end{aligned} \qquad (1 \le t \le n), \tag{26}$$

де $U_t f(a) = q_t(a) + \sum\limits_{y \in X_t} p_t(y\,|a) f(y) + p_t(y^*\,|a) c(y^*)\,,\ (a \in A_t,\ y^* \in X_t)$,
$V_t g(x) = \sup\limits_{A(x)} g(a),\ (x \in X_{t-1})$, причому $\upsilon_n = r$.

Рівності (26) називаються рівняннями оптимальності. Приймемо $T_t = V_t U_t$, тоді рівняння оптимальності запишемо у вигляді

$$\upsilon_{t-1} = T_t \upsilon_t. \tag{27}$$

Теорема 5. *Нехай π – довільна обривна стратегія в похідній моделі Z_k^* $(k = 1, 2, ...n)$ і ψ_t – довільні селектори багатозначного відображення $A(x)$: $X_{t-1} \to A_t$ $\quad (t = 1, 2, ...k)$, тоді:*

$$\omega_0(x, \psi_1\psi_2...\psi_k\pi) = T_{\psi_1} T_{\psi_2} ... T_{\psi_k} \omega(x, \pi). \tag{28}$$

Доведення. Випливає з фундаментального рівняння (10), формул (12), (27) і методу математичної індукції. □

3. Обривні керовані марковські процеси для зліченних моделей. Для зліченних моделей оптимальні стратегії можуть не існувати, тому вводиться поняття ε-оптимальності.

Означення 3. *Обривна стратегія π називаєтсья ε-оптимальною для процесу Z^*_μ, якщо $\omega(\mu,\pi) \geq \upsilon(\mu) - \varepsilon$.*

Означення 4. *Обривна стратегія π називається рівномірно ε-оптимальною або ε-оптимальною для моделі Z^*, якщо π-ε-оптимальна для Z^*_μ і будь-якого μ – початкового розподілу.*

Далі з'ясуємо, як результати для скінченних моделей з оптимальними стратегіями переносяться на зліченні моделі з ε-оптимальними стратегіями.

Нехай π_x-ε-оптимальна стратегія для процесу Z^*_x. Вона існує за означенням верхньої грані.

За набором обривних стратегій π_x знову хочемо побудувати одну обривну стратегію π, ε-оптимальну для моделі.

Доведемо, що формула (6) визначає деяку стратегію $\overline{\pi}$, яка буде ε-оптимальною, тобто $\omega(x,\overline{\pi}) = \omega(x,\pi_x) \geq \upsilon(x) - \varepsilon,\ \forall x \in X_m$.

Теорема 6. *Будь-яка обривна стратегія $\overline{\pi}$, задана формулою (6), для якої*

$$\omega(x,\overline{\pi}) \geq \upsilon(x) - \varepsilon,\ (x \in X_m) \tag{29}$$

рівномірно ε-оптимальна, тобто $\forall\mu\ \sup\limits_{\pi}\omega(\mu,\pi) \leq \omega(\mu,\overline{\pi}) + \varepsilon$.

Доведення. З формул (2)–(4) випливає, що $\forall\pi$

$$\omega(\mu,\pi) = \sum_{l\in L} I(l)P^*(l) = \sum_{X_m}\mu(x)\omega(x,\pi). \tag{30}$$

Звідси отримуємо

$$\omega(\mu,\pi) = \sum_{X_m}\mu(x)\omega(x,\pi) \leq \sum_{X_m}\mu(x)\upsilon(x) \leq \sum_{X_m}\mu(x)[\omega(x,\overline{\pi})+\varepsilon] = \omega(\mu,\overline{\pi}) + \varepsilon. \tag{31}$$

З отриманої нерівності випливає:

$$\sup_{\pi}\omega(\mu,\pi) \leq \sum_{X_m}\mu(x)\upsilon(x), \tag{32}$$

$$\omega(\mu,\overline{\pi}) \geq \sum_{X_m}\mu(x)\upsilon(x) - \varepsilon. \tag{33}$$

З формул (32) і (33) отримуємо

$$\sup_{\pi}\omega(\mu,\pi) \leq \sum_{X_m}\mu(x)\upsilon(x) \leq \omega(\mu,\overline{\pi}) + \varepsilon. \tag{34}$$

Отже, стратегія $\overline{\pi}$ рівномірно ε-оптимальна. Оскільки це доведення можемо провести для $\forall\varepsilon$, то $\sup\limits_{\pi}\omega(\mu,\pi) = \sum\limits_{X_m}\mu(x)\upsilon(x) = \mu\upsilon$. □

Наслідок 5. *Для довільного початкового розподілу μ*

$$\upsilon(\mu) = \mu\upsilon. \tag{35}$$

Зауваження 2. Зауважимо, що фундаментальне рівняння правильне і для зліченних моделей без змін у доведенні.

Теорема 7. *Нехай $\pi = \gamma\pi'$ – добуток обривних стратегій γ та π'. Якщо π' рівномірно ε'-оптимальна для моделі Z'^*, тоді*

$$\upsilon = Vu. \tag{36}$$

Доведення. Для добутку стратегій $\gamma\pi'$ фундаментальне рівняння (10) набуде такого вигляду:

$$\omega(x, \gamma\pi') = \sum_{A(x)} \gamma(a\,|x)(q(a) + \omega'(p_a, \pi')). \tag{37}$$

Оскільки π' – рівномірно ε'-оптимальна (а вона існує $\forall\varepsilon' > 0$ згідно з теоремою 1), то $\omega'(p_a, \pi') \geq \upsilon'(p_a) - \varepsilon'$ і згідно з виглядом функції u рівняння (37) перетворюється в

$$\omega(x, \gamma\pi'0 \geq \sum_{A(x)} \gamma(a\,|x)u(a) - \varepsilon'. \tag{38}$$

Розглянемо множину $A_\chi(x) = \{a: \ a \in A(x), \ u(a) \geq Vu(x) - \chi\} \quad (x \in X_m)$, яка не порожня для будь-якого $\chi > 0$. Нехай $\gamma(\cdot\,|x)$ довільний розподіл ймовірностей на $A(x)$, зосереджений на $A_\chi(x)$. Тоді $\sum\limits_{A(x)} \gamma(a\,|x)u(a) \geq Vu(x) - \chi$.

При $\varepsilon' + \chi \leq \varepsilon$

$$\omega(x, \pi) \geq Vu(x) - \varepsilon \quad (x \in X_m). \tag{39}$$

□

Наслідок 6. *Оцінка υ моделі Z^* виражається через оцінку υ' моделі Z'^* такими формулами:*

$$\begin{aligned} \upsilon &= Vu, \\ u &= U\upsilon', \end{aligned} \tag{40}$$

де оператори U та V задані формулами (18) та (19).

Наслідок 7. *Для будь-якого $\chi > 0$ існує селектор ψ багатозначного відображення $A(x): \ X_m \to A_{m+1}$*

$$u(\psi(x)) \geq \upsilon(x) - \chi. \tag{41}$$

За $\gamma(\cdot\,|x)$ можна прийняти розподіл, зосереджений в одній точці $\psi(x) \in A_\chi(x)$.

Наслідок 8. *Нехай ε' і χ – довільні невід'ємні числа. Якщо π' рівномірно ε'-оптимальна для моделі Z'^* і селектор ψ – такий, як у наслідку 7, то обривна стратегія $\psi\pi'$ – рівномірно $(\varepsilon' + \chi)$-оптимальна для моделі Z^*.*

Теорема 8. *Для фіксованого початкового розподілу μ та для довільної обривної стратегії π існує проста стратегія φ*

$$\omega(\mu, \pi) \leq \omega(\mu, \varphi). \tag{42}$$

Доведення цієї теореми випливає з таких двох теорем.

Теорема 9. *$\forall\mu$ і для будь-якої обривної стратегії π існує марковська стратегія σ така що:*

$$\omega(\mu, \sigma) = \omega(\mu, \pi). \tag{43}$$

Доведення. Розглянемо марковську стратегію σ

$$\sigma(a\,|x) = \mathbb{P}^*\{a_t = a \mid x_{t-1} = x\} = \frac{\mathbb{P}^*\{x_{t-1}a_t = xa\}}{\mathbb{P}^*\{x_{t-1} = x\}} \tag{44}$$

$$(a \in A_t, \quad x \in X_{t-1}, \quad m+1 \le t \le n),$$

де $\mathbb{P}^*$ – міра в просторі шляхів L, яка відповідає початковому розподілу μ та стратегії π.

Якщо у правій частині формули (44) при $\mathbb{P}^*\{x_{t-1} = x\} = 0$ тоді вираз втрачає зміст. Для таких x(зокрема обривних) за $\sigma(\cdot\,|x)$ можна вибрати довільний розподіл на $A(x)$.

Позначимо через $\mathbb{Q}^*$ – розподіл ймовірностей у просторі L, що відповідає початковому розподілу μ та обривній марковській стратегії σ. В загальному випадку розподіл $\mathbb{Q}^*$ не збігається з розподілом $\mathbb{P}^*$, але для доведення (43) достатньо, щоб кожен із елементів $x_m, ..., x_n, a_{m+1}, .., a_n$ та $x^*_{m+1}, ...x^*_n$ мав однаковий розподіл ймовірностей стосовно мір $\mathbb{P}^*$ та $\mathbb{Q}^*$. Це випливає з таких двох рівностей:

$$\omega(\mu,\pi) = \sum_{t=m+1}^{n} \mathbb{P}^* q(a_t) + \sum_{t=m+1}^{n} \mathbb{P}^* c(x^*_t) + \mathbb{P}^* r(x_n) \tag{45}$$

$$\omega(\mu,\sigma) = \sum_{t=m+1}^{n} \mathbb{Q}^* q(a_t) + \sum_{t=m+1}^{n} \mathbb{Q}^* c(x^*_t) + \mathbb{Q}^* r(x_n). \tag{46}$$

Доводити будемо за індукцією.

Твердження правильне для x_m, бо розподіл для нього

$$\mathbb{P}^* = \mathbb{Q}^* = \mu.$$

Припустимо, що твердження виконується для x_{t-1}. Перевіримо це для a_t. Оскільки обривна стратегія σ – марковська, то

$$\mathbb{Q}^*\{x_{t-1}a_t = xa\} = \mathbb{Q}^*\{x_{t-1} = x\}\sigma(a\,|x), \quad (a \in A_t, \quad x \in X_{t-1}). \tag{47}$$

Тоді з (44) та (47) отримуємо таке:

$$\begin{aligned} \mathbb{P}^*\{a_t = a\} &= \sum_{x\in X_{t-1}} \mathbb{P}^*\{x_{t-1}a_t = xa\} = \sum_{x\in X_{t-1}} \mathbb{P}^*\{x_{t-1} = x\}\sigma(a\,|x) = \\ &= \sum_{x\in X_{t-1}} \mathbb{Q}^*\{x_{t-1} = x\}\sigma(a\,|x) = \sum_{x\in X_{t-1}} \mathbb{Q}^*\{x_{t-1}a_t = xa\} = \mathbb{Q}^*\{a_t = a\}. \end{aligned} \tag{48}$$

Отже, наше твердження правильне для a_t. Доведемо тепер таке: якщо воно правильне для a_t, то й виконується для x_t.

За означенням перехідної функції

$$\mathbb{P}^*\{a_t x_t = ax\} = \mathbb{P}^*\{a_t = a\}p(x\,|a), \tag{49}$$

$$\mathbb{Q}^*\{atxt = ax\} = \mathbb{Q}^*\{at = a\}p(x\,|a). \tag{50}$$

Відразу з (49) і (50) отримуємо

$$\begin{aligned} \mathbb{P}^*\{x_t = x\} &= \sum_{x\in X_{t-1}} \mathbb{P}^*\{a_t x_t = ax\} = \sum_{x\in X_{t-1}} \mathbb{P}^*\{a_t = a\}p(x\,|a) = \\ &= \sum_{x\in X_{t-1}} \mathbb{Q}^*\{a_t = a\}p(x\,|a) = \sum_{x\in X_{t-1}} \mathbb{Q}^*\{a_t x_t = ax\} = \mathbb{Q}^*\{x_t = x\}. \end{aligned} \tag{51}$$

□

Лема 1. *Нехай f – довільна функція і υ – довільний розподіл ймовірностей на зліченному просторі E. Якщо $\upsilon f < +\infty$, то множина $\Gamma = \{x : \ f(x) \geq \upsilon f\}$ має додатну міру υ, тобто $\upsilon(\Gamma) > 0$.*

Теорема 10. *Для будь-якої марковської стратегії σ існує проста стратегія φ така що:*

$$\omega(\mu, \varphi) \geq \omega(\mu, \sigma). \tag{52}$$

Доведення. Згідно з (30) умова (52) рівносильна

$$\omega(x, \varphi) \geq \omega(x, \sigma), \quad \forall x \in X_m. \tag{53}$$

Розкладемо обривну марковську стратегію σ в добуток стратегій $\sigma = \gamma\sigma'$, де γ – звуження σ на X_m, а σ' – звуження σ в на $X_{m+1} \bigcup X_{m+2} ... \bigcup X_n$. Згідно з фундаментальним рівнянням (10)

$$\omega(x, \sigma) = \gamma_x f, \tag{54}$$

де $\gamma_x(\cdot) = \gamma(\cdot\,|x)$ – розподіл ймовірностей на $A(x)$, і $f(a) = q(a) + \omega'(p_a, \sigma')$, $(a \in A_{m+1})$.

За лемою 1 підмножина $A(x)$, для якої $f(a) \geq \gamma_x f = \omega(x, \sigma)$, має додатню міру γ_x, а отже, є непорожньою. Якщо $\psi(x)$ – довільна точка цієї підмножини, то $f(\psi(x)) \geq \omega(x, \sigma)$. На підставі фундаментального рівняння (10), $f(\psi(x)) = \omega(x, \psi\sigma')$, а отже, $\omega(x, \psi\sigma') \geq \omega(x, \sigma)$.

Припустимо, що умова (52) правильна для похідної моделі Z'^*, тоді в цій моделі знайдеться проста стратегія φ', яка рівномірно мажорує обривну марковську стратегію σ'. З фундаментального рівняння (10) та зробленого припущення отримуємо

$$\omega(x, \psi\varphi') = q(\psi(x)) + \omega'(p_{\psi(x)}, \varphi') \geq q(\psi(x)) + \omega'(p_{\psi(x)}, \sigma') = \omega(x, \psi\sigma') \geq \omega(x, \sigma). \tag{55}$$

Отже, в моделі Z^* проста стратегія $\varphi = \psi\,\varphi'$ рівномірно мажорує σ так, що результат (52) є правильний і для моделі Z^*. $\square$

Список використаної літератури

1. *Дынкин Е.Б.* Управляемые марковские процессы и их приложения /*Е.Б. Дынкин, А.А. Юшкевич* – М.: Наука, 1975. – 334 с.
2. *Пароля Н.Р.* Обривні керовані марковські процеси на скінченному інтервалі часу для скінченних моделей / *Н.Р. Пароля, Я.І. Єлейко* // Вісн. Львів. ун-ту. Сер. мех.-мат. – 2010. – № 72 – С. 243–254.

Стаття: надійшла до редакції 07.10.2013
прийнята до друку 16.10.2013

KILLED MARKOV DECISION PROCESSES WITH FINITE OR COUNTABLE SETS OF STATES AND MANAGEMENTS

Pavlo SHPAK, Yaroslav YELEYKO

Ivan Franko National University of Lviv,
Universytetska Str., 1, Lviv, 79000
e-mail: prshpak@gmail.com

We present new approach to probability measure on the space of paths, redefined concept of paths and crash function. Correctness of definitions and existence of optimal, uniformly optimal, ε-optimal and uniformly ε-optimal strategies under the new probability measure is shown. Sufficient of simple strategies in finite and countable models is proved.

Key words: killed Markov desision process, optimal strategy, ε-optimal strategy, uniformly ε-optimal strategy.

ОБРЫВАЮЩИЕСЯ УПРАВЛЯЕМЫЕ МАРКОВСКИЕ ПРОЦЕССЫ С СЧЕТНЫМИ МНОЖЕСТВАМИ СОСТОЯНИЙ И УПРАВЛЕНИЙ

Павло ШПАК, Ярослав ЕЛЕЙКО

Львовский национальный университет имени Ивана Франко,
ул. Университетская, 1, Львов, 79000
e-mail: prshpak@gmail.com

Предложен новый подход к введению вероятностной меры на пространстве путей, изменено определение пути в рамках обрывающегося управляемого марковского процесса и функцию краха. Показано корректность определений и существование оптимальных, равномерно оптимальных, ε-оптимальных и равномерно ε-оптимальных стратегий в условиях новой вероятностной меры. Доказано достаточность простых стратегий в конечной и счетной моделях.

Ключевые слова: обрывающиейся управляемый марковский процесс, оптимальная стратегия, ε-оптимальная стратегия, равномерно ε-оптимальная стратегия.