ВІСНИК ЛЬВІВ. УН-ТУ
Серія мех.-мат. 2010.
Вип. 72. С. 243–254

VISNYK LVIV UNIV.
Ser. Mech. Math. 2010.
Is. 72. P. 243–254

УДК 519.2:519.7

ОБРИВНІ КЕРОВАНІ МАРКОВСЬКІ ПРОЦЕСИ НА СКІНЧЕННОМУ ІНТЕРВАЛІ ЧАСУ ДЛЯ СКІНЧЕННИХ МОДЕЛЕЙ

Нестор ПАРОЛЯ, Ярослав ЄЛЕЙКО

Львівський національний університет імені Івана Франка,
79000 Львів, вул. Університетська, 1
e-mail: mstat@franko.lviv.ua

Розглянуто обривні керовані марковські процеси для скінченних моделей на скінченному інтервалі часу. Доведено існування рівномірно оптимальної стратегії. Показано правильність фундаментального рівняння. Зведено задачу оптимального керування до аналогічної задачі для похідної моделі. Виведено рівняння оптимальності та метод побудови простих рівномірно оптимальних стратегій. Показано правильність марковської властивості. Розглянуто принцип динамічного програмування.

Ключові слова: керований марковський процес, оптимальна стратегія, оптимальне керування, фундаментальне рівняння.

1. Вступ. Керовані марковські процеси виникають у найрізноманітніших сферах економіки, зокрема для економічного планування роботи окремого підприємства, галузі або всього народного господарства. На початку кожного періоду, враховуючи досягнуте, можна побудувати план на наступний період. Розвиток системи можна описати математично керованим детермінованим процесом, якщо вважати, що стан системи в кінці кожного періоду однозначно визначається станом на початку періоду і планом на цей період. Проте не можна нехтувати впливом таких чинників: метеорологічні умови, демографічні зсуви, коливання попиту, недосконалість координації складних виробничих процесів, наукові відкриття, винаходи тощо. Ці чинники ліпше враховують стохастичні моделі, в яких, знаючи стан на початку періоду і план, можна обчислити лише розподіл ймовірності для стану в кінці періоду. Якщо ж не враховувати стан системи в минулих періодах, то ми переходимо до керованого марковського процесу ("майбутнє залежить від теперішнього і не залежить від минулого").

Процеси такого типу детально описано в [1]: було введено означення керованого марковського процесу, поняття "марковської моделі" Z^μ, стратегії π, оцінки стратегії $\omega(\pi)$, ν-оцінки процесу Z^μ, доведено існування рівномірно оптимальної стратегії, виведено фундаментальне рівняння, рівняння оптимальності та схему знаходження простих оптимальних стратегій, доведено "марковську властивість" та показано принцип динамічного програмування.

У згаданій моделі не враховано фактор ризику, а саме ймовірність того, що підприємство може зазнати краху, збанкрутувати в певний момент часу. Іншими словами, ми приходимо до поняття обривного керованого марковського процесу, де в кожний момент часу, крім початкового, звісно, процес може обірватися з певною ненульовою ймовірністю.

Принцип обривних марковських моделей ще більше наближає стохастичну модель до реальності, яка неможлива без ризику.

2. Означення обривного керованого марковського процесу. Нехай $X_t(t = m, \dots, n)$ та $A_t(t = m+1, \dots, n)$ довільні скінченні множини. $\forall a \in A_t$ ставиться у відповідність розподіл ймовірностей $p(\cdot|a)$ на X_t.

Означення 1. *Функцію p, що визначає закон переходу з A_t в X_t, називатимемо* ***перехідною функцією****.*

Означення 2. *Точка $x^* = x_m \in X_t$ називається* ***точкою обриву****, а $p(x^*|a)$ –* ***ймовірністю обриву****, якщо $\mathbb{P}(x_{t+1} = x^*|a_t = a) = \mathbb{P}(x_{t+1} = x_m|a_t = a) \equiv p(x^*|a)$, $x_m \in X_m$.*

Зауваження 1. Іншими словами, потрапивши в точку обриву, система переходить в початковий стан (процес обривається).

З означення обривної точки випливає

$$\forall\, a \in A_t \quad \exists\, x^* \in X_t : \quad p(x^*|a) = 1 - \sum_{x \in X_t \setminus x^*} p(x|a) > 0.$$

Означення 3 (Обривного керованого марковського процесу [2])**.** *Обривним керованим марковським процесом на інтервалі часу $[m, n]$ називається випадковий процес, який задається такими об'єктами.*

1. Множини $X_m, \dots, X_n$ (простори станів).

2. Множини $A_{m+1}, \dots, A_n$ (простори керувань).

3. Відображення проектування $j : A \to X$, де $A = \bigcup\limits_{t=m+1}^{n} A_t$, $X = \bigcup\limits_{t=m}^{n} X_t$: $j(A_t) = X_{t-1} \setminus \{x^*\}$, $x^* \in X_{t-1}$, $(t = m+2, \dots, n)$ і $j(A_{m+1}) = X_m$.

4. Розподіл ймовірностей $p(\cdot|a)$ на X_t з обривними точками

$$\mathbb{P}(x_{t+1} = x^*|a_t = a) = \mathbb{P}(x_{t+1} = x_m|a_t = a) \equiv p(x^*|a) > 0.$$

5. Функція q, задана на A (функція винагороди).

6. Функція r, задана на множині X_n (фінальна плата).

7. Функція c (від англ. "crash" – крах, розорення), задана в обривних точках станів $c(x^*) = -\sum\limits_{i=m+1}^{t} \max\limits_{a_i \in A_i} q(a_i)$, $x^* \in X_t$, $t = m+1, \dots, n$ (функція краху, задана

так, гарантує повне банкрутство – повну втрату накопиченого капіталу або й більше).

8. Початковий розподіл μ на X_m.

*Процес, що задовольняє (1-8), називатимемо ([3]) **обривним марковським процесом (моделлю)** і позначатимемо Z^*_μ. Якщо початковий розподіл μ зосереджений в точці x, то будемо писати Z^*_x.*

Наша мета – знайти спосіб керування, за якого максимізується математичне сподівання оцінки шляху l

$$I(l, x^*) = \sum_{t=m+1}^{n} [q(a_t) + c(x^*_t)] + r(x_n), \tag{1}$$

де $x^* = (x^*_{m+1}, \ldots, x^*_n)$ – вектор точок обриву; $l = x_m a_{m+1}, \ldots, a_n x_n$ – шлях.

Під способом керування будемо розуміти певну *стратегію*.

3. Стратегії.

Означення 4. *Селектор багатозначного відображення $A(x) : X \setminus X_n \to A$ $\varphi(x) : A(x) \to A$ називається **простою стратегією**, якщо $\varphi(x_{t-1}) = a_t$, для довільного x_t – необривних точок станів з ймовірнісними розподілами $p(\cdot|a_t)(m < t \leqslant n)$ та x_m з початковим розподілом μ.*

Зауваження 2. $\varphi(x)$ не визначена в обривних точках. Крім того, користуючись простою стратегією $\varphi(x)$ ми одержуємо шлях $l = x_m a_{m+1}, \ldots, a_n x_n$.

Означення 5. *Відображення $\pi : H \to \pi(\cdot|h \in H)$, де $\pi(\cdot|h \in H)$ – розподіл ймовірностей на $A(x_{t-1})$ і H – множина історій ($h \in H$ тоді і тільки тоді, коли $h = x_m a_{m+1}, \ldots, a_{t-1} x_{t-1}$), називається **обривною стратегією**.*

Зауваження 3. $x_{t-1} \neq x^*$.

Означення 6. *Обривна стратегія $\pi(\cdot|h)$ називається **марковською**, якщо $\pi(\cdot|h) = \pi(\cdot|x_{t-1})$.*

Позначимо L – множину всіх шляхів l.

Означення 7. *Якщо задані перехідна функція $p(\cdot|a)$ і стратегії $\pi(\cdot|h)$, то кожному початковому розподілу μ відповідає розподіл ймовірностей P^* у просторі L, який набуває такого вигляду:*

$$P^*(l, x^*) = P^*(x_m a_{m+1}, \ldots, a_n x_n,\ x^*_{m+1}, \ldots, x^*_n) =$$

$$= \mu(x_m)\pi(a_{m+1}|x_m)p(x_{m+1}|a_{m+1})p(x^*_{m+1}|a_{m+1})\cdot\ldots\cdot\pi(a_n|h_{n-1})p(x_n|a_n)p(x^*_n|a_n). \tag{2}$$

Зауваження 4. Після того, як ми означили міру P^*, шлях l можна вважати випадковим процесом. Крім того, якщо стратегія π – марковська, то цей процес є марковським.

Для будь-якої функції $\xi = \xi(l)$ з простору L математичне сподівання ξ виглядатиме так:

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

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

$$\omega = E^* I(l, x^*) = E^*[\sum_{t=m+1}^{n} [q(a_t) + c(x_t^*)] + r(x_n)]. \tag{4}$$

Означення 8 (Оцінки стратегії). *Величину ω з (4) для обривного керованого процесу Z_μ^*, яка є функцією від π ($\omega = \omega(\pi)$), називатимемо* ***оцінкою стратегії*** *π.*

Мета досліджень – максимізація функції $\omega(\pi)$.

Означення 9 (Оцінки обривного процесу). *$\nu \equiv \sup\limits_{\pi} \omega(\pi)$ називається* ***оцінкою обривного процесу*** *Z_μ^* або* ***оцінкою початкового розподілу*** *μ.*

Зауваження 5. $\nu(x^*) = c(x^*)$.

Означення 10 (Оптимальної стратегії). *Обривна стратегія π називається* ***оптимальною****, якщо $\omega(\pi) = \nu$.*

Означення 11 (Рівномірно оптимальна стратегія). *Обривна стратегія π називається* ***рівномірно оптимальною*** *або оптимальною для процесу Z^*, якщо π – оптимальна для Z_μ^* для кожного μ – початкового розподілу.*

Твердження 1 (Про змішування стратегій). *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 ми будемо використовувати обривну стратегію π_k з ймовірністю γ_k, то одержимо 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 деяка обривна стратегія π, якій відповідає розподіл ймовірностей P^* з (5).

Доведення. Приймемо

$$\pi(a_{t+1}|x_m a_{m+1} \dots x_t) = \tag{6}$$

$$= \begin{cases} \dfrac{\sum\limits_k \gamma_k \pi_k(a_{m+1}|x_m) \dots \pi_k(a_t|h_{t-1}) \pi_k(a_{t+1}|h_t)}{\sum\limits_k \gamma_k \pi_k(a_{m+1}|x_m) \dots \pi_k(a_t|h_{t-1})}, & \text{якщо знаменник не нуль,} \\ \pi_1(a_{t+1}|h_t), & \text{в іншому випадку;} \end{cases}$$

тут $m \leqslant t < n$, $h_t = x_m a_{m+1} \dots x_t$ – довільна історія, a_{t+1} – довільне керування з A_t; при $t = m$ вважається, що знаменник дорівнює 1.

Вигляд $\pi(a_{t+1}|x_m a_{m+1} \dots x_t)$ випливає з (5) та такого співвідношення:

$$\pi(a_{t+1}|x_m a_{m+1} \dots x_t) = \frac{P^*(x_m a_{m+1} \dots a_t x_t a_{t+1})}{P^*(x_m a_{m+1} \dots a_t x_t)}.$$

З того, що $\pi_k(\cdot|x_m a_{m+1}\dots x_t)$ – розподіл ймовірностей, сконцентрований на $A(x_t)$, та умови $\sum\limits_k \gamma_k = 1, \gamma_k \geqslant 0$, випливає, що $\pi(\cdot|x_m a_{m+1}\dots x_t)$ – теж розподіл ймовірностей, сконцентрований на $A(x_t)$. А отже, $\pi(\cdot|x_m a_{m+1}\dots x_t)$, зображена формулою (6), – стратегія.

Тепер з формули (6) $\forall l = x_m a_{m+1}\dots x_n$ одержимо

$$\pi(a_{m+1}|x_m)\pi(a_{m+2}|x_m a_{m+1} x_{m+1})\cdot\ldots\cdot\pi(a_n|x_m a_{m+1}\dots x_{n-1}) =$$

$$= \sum_k \gamma_k \pi_k(a_{m+1}|x_m)\pi_k(a_{m+2}|x_m a_{m+1} x_{m+1})\cdot\ldots\cdot\pi_k(a_n|x_m a_{m+1}\dots x_{n-1}).$$

Домножуємо праву та ліву частину рівності на

$$\mu(x_m)p(x_{m+1}|a_{m+1})p^*(x^*_{m+1}|a_{m+1})\cdot\ldots\cdot p(x_n|a_n)p^*(x^*_n|a_n)$$

і враховуючи формулу (2), отримуємо, що обривній стратегії π відповідає міра $\sum\limits_k \gamma_k P^*_k = P^*$ □

Зауваження 6. 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.

4. Існування рівномірно оптимальної стратегії. Поєднання стратегій. Обривна стратегія π описується скінченним набором невід'ємних чисел $\pi(a|h)$. Набори, які задають стратегію, утворюють замкнуту обмежену множину Π в скінченновимірному просторі. А отже, Π – компакт. Функція $\omega(\pi)$ неперервна, бо виражається через $\pi(a|h)$ за допомогою операцій множення та додавання. За теоремою Вейєрштрасса, неперервна функція на Π досягає свого максимуму. Та обривна стратегія, за якої 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 ми хочемо побудувати одну обривну стратегію π, оптимальну для процесу Z^*.

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

$$\bar{\pi}(\cdot|h) = \pi_{x(h)}(\cdot|h), \tag{7}$$

де $x(h)$ – початковий стан історії h. Звісно, що формула (7) визначає деяку стратегію $\bar{\pi}$, яка буде оптимальною, тобто $\omega(x,\bar{\pi}) = \omega(x,\pi_x) = \nu(x), \forall x \in X_m$.

Твердження 2 (Про існування рівномірно оптимальної обривної стратегії)**.** *Будь-яка обривна стратегія $\bar{\pi}$, задана формулою (7),для якої*

$$\omega(x,\bar{\pi}) = \nu(x), (x \in X_m)$$

є рівномірно оптимальною, тобто для кожного μ: $\sup\limits_\pi \omega(\mu,\pi) = \omega(\mu,\bar{\pi})$.

Доведення. З формул (2)-(4) випливає, що для кожного π

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

Зокрема, $\omega(\mu,\bar{\pi}) = \sum\limits_{X_m} \mu(x)\omega(x,\bar{\pi})$. Але $\omega(x,\pi) \leqslant \omega(x,\bar{\pi})$ для всіх $x \in X_m$, а отже, $\omega(\mu,\pi) \leqslant \omega(\mu,\bar{\pi})$. □

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

$$\nu(\mu) = \mu\nu. \tag{9}$$

Доведення. Випливає з рівності

$$\nu(\mu) = \omega(\mu,\bar{\pi}) = \sum_{X_m} \mu(x)\omega(x,\bar{\pi}) = \sum_{X_m} \mu(x)\nu(x) = \mu\nu.$$

□

Зауваження 7. Формули (8) та (9) допомагають звести вивчення процесів Z^*_μ при довільному μ до вивчення процесів $Z^*_x, \forall x \in X_m$.

Стратегія π, побудована за набором π_x ($x \in X_m$), має таку властивість (*):

Для будь-якого початкового стану $x \in X_m$ розподіли ймовірностей у просторі шляхів L, які відповідають за формулою (2) стратегіям π та π_x, ідентичні.

Означення 12. *Якщо виконується властивість (*), то стратегія $\bar{\pi}$ називається* ***поєднанням*** *стратегій π_x.*

5. Похідна модель. Фундаментальне рівняння. Процес керування – це низка послідовних кроків. Перший крок полягає у виборі розподілу ймовірностей на A_{m+1}(що залежить від початкового стану). Якщо вибір зроблений, то кожному початковому розподілу μ на X_m відповідає розподіл ймовірностей $\acute{\mu}$ на X_{m+1}. Тепер розглядаємо $\acute{\mu}$ як початковий розподіл в момент часу $m+1$.

Розбиваємо нашу задачу максимізації на дві.

1. За будь-якого початкового розподілу на X_{m+1} вибрати оптимальну поведінку в подальші моменти часу.

2. Вибрати перший крок так, щоб була максимальна сума винагороди за цей крок і максимальна величина оцінки оптимальної поведінки в подальші моменти при початковому розподілі $\acute{\mu}$.

Означення 13 (Похідної моделі)**.** *Обривна модель, яка утворена з моделі Z^* викресленням X_m та A_{m+1}, називається похідною моделлю і позначається $\acute{Z}^*$.*

Твердження 3 (Фундаментальне рівняння)**.**

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

де $p_a = p(\cdot|a), \pi_a(\cdot|\acute{h}) = \pi(\cdot|ya\acute{h})$, $a \in A_{m+1}, y = j(a)$, $\acute{h}$ – історія в моделі $\acute{Z}^*$.

Рівняння (10) називається ***фундаментальним*** *і виражає оцінку ω довільної стратегії π в моделі Z^* через оцінку $\acute{\omega}$ деяких стратегій в моделі $\acute{Z}^*$*

Доведення. Згідно з формулою (8) отримуємо

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

Розглянемо простори шляхів L та $\acute{L}$ в моделях Z^* та $\acute{Z}^*$. Нехай P^* – розподіл ймовірностей в L, який відповідає початковому стану x та стратегії π, P_a^* – розподіл ймовірностей в $\acute{L}$, який відповідає початковому розподілу p_a та стратегії π_a.

Згідно з формулою (1) та (2) $\forall \acute{l} \in \acute{L}$ одержуємо

$$I(xa\acute{l}, x^*) = q(a) + I(\acute{l}, x^*_{-1}) \tag{12}$$

$$P^*(xa\acute{l}, x^*) = \pi(a|x) P_a^*(\acute{l}, x^*_{-1}) \tag{13}$$

$$a \in A(x), \quad x^*_{-1} = (x^*_{m+2}, \ldots, x^*_n), \quad (x^*_{m+1}, x^*_{-1}) = x^*.$$

На підставі (3) і (4) отримуємо

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

$$\acute{\omega}(p_a, \pi_a) = \sum_{\acute{L}} P_a^*(\acute{l}, x^*_{-1}) I(\acute{l}, x^*_{-1}). \tag{15}$$

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

6. Зведення задачі оптимального керування до аналогічної задачі для похідної моделі. З фундаментального рівняння (10) випливає така оцінка:

$$\omega(x, \pi) \leqslant \sup_{A(x)} [q(a) + \acute{\omega}(p_a, \pi_a)] \leqslant \sup_{A(x)} [q(a) + \acute{\nu}(p_a)] \tag{16}$$

$\forall x \in X_m$ і $\forall \pi$ ($\acute{\nu}$ – оцінка моделі $\acute{Z}^*$).

Позначимо $u(a) = q(a) + \acute{\nu}(p_a), (a \in A_{m+1})$. Будемо називати цю величину – **оцінкою керування** a.

Згідно з (9) та рівністю $\nu(x^*) = c(x^*)$ отримуємо $u = U\acute{\nu}$, де оператор U переводить функції на точках (необривних) у функції на керуваннях за такою формулою:

$$Uf(a) = q(a) + \sum_y p(y|a) f(y) + \sum_{y^*} p(y^*|a) c(y^*), \tag{17}$$

де y – необривні точки; y^* – обривні точки.

Введемо ще оператор V, який переводить функції на керуваннях у функції на точках(нефінальних і необривних) так:

$$Vg(x) = \sup_{a \in A(x)} g(a). \tag{18}$$

Запишемо нерівність (16), використовуючи оператор V: $\omega(x, \pi) \leqslant Vu(x)$. Тепер візьмемо $\sup\limits_{\pi}$ від лівої та правої частини й отримаємо $\nu \leqslant Vu$. Пізніше ми покажемо, за яких умов справджується рівність.

Означення 14 (Добутку стратегій)**.** *Нехай $\acute{\pi}$ – довільна обривна стратегія в моделі $\acute{Z}^*$ і нехай $\forall x \in X_m$ поставлено у відповідність деякий розподіл ймовірностей $\gamma(\cdot|x)$ на A_{m+1}, який зосереджений на $A(x)$. Вибираючи на початковому кроці керування a з розподілом γ і користуючись на всіх наступних кроках обривною стратегією*

$\acute{\pi}$, ми отримаємо обривну стратегію π в моделі Z^*, яка називається ***добутком*** *γ та $\acute{\pi}$, позначається $\gamma\acute{\pi}$ і описується формулою*

$$\pi(\cdot|h) = \begin{cases} \gamma(\cdot|x) & \text{при } h = x \in X_m, \\ \acute{\pi}(\cdot|\acute{h}) & \text{при } h = xa\acute{h}. \end{cases}$$

Твердження 4. *Нехай $\pi = \gamma\acute{\pi}$ – добуток обривних стратегій γ та $\acute{\pi}$. Якщо $\acute{\pi}$ рівномірно оптимальна для моделі $\acute{Z}^*$, тоді*

$$\nu = Vu. \tag{19}$$

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

$$\omega(x, \gamma\acute{\pi}) = \sum_{A(x)} \gamma(a|x)\Big(q(a) + \acute{\omega}(p_a, \acute{\pi})\Big). \tag{20}$$

Оскільки $\acute{\pi}$ – рівномірно оптимальна(а вона існує згідно з *твердженням 2*), то $\acute{\omega}(p_a, \acute{\pi}) = \acute{\nu}(p_a)$, і згідно з виглядом функції u рівняння (20) перетворюється в

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

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

$$\omega(x, \gamma\acute{\pi}) = Vu(x) \ (x \in X_m). \tag{21}$$

□

Наслідок 2. *Оцінка ν моделі Z^* виражається через оцінку $\acute{\nu}$ моделі $\acute{Z}^*$ такими формулами:*

$$\nu = Vu, \ u = U\acute{\nu}, \tag{22}$$

де оператори U та V задані формулами (17) та (18), відповідно.

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

$$u(\psi(x)) = \nu(x). \tag{23}$$

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

□

Наслідок 4. *Якщо $\acute{\pi}$ рівномірно оптимальна для моделі $\acute{Z}^*$ і селектор ψ – такий, як в Наслідку 2, то обривна стратегія $\psi\acute{\pi}$ – рівномірно оптимальна для процесу Z^*.*

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

Згідно з результатами попереднього параграфа оцінки ν_t і u_t пов'язані між собою такими співвідношеннями:

$$\nu_{t-1} = Vu_t, \quad u_t = U\nu_t \quad (1 \leqslant t \leqslant n), \tag{24}$$

де

$$U_t f(a) = q_t(a) + \sum_{y \in X_t} p_t(y|a) f(y) + p_t(y^*|a) c(y^*), \quad (a \in A_t, \quad y^* \in X_t),$$

$$V_t g(x) = \sup_{A(x)} g(a), \quad (x \in X_{t-1}),$$

причому $\nu_n = r$.

Рівності (24) називаються *рівняннями оптимальності*. Приймемо $T_t = V_t U_t$, тоді рівняння оптимальності запишемо у вигляді

$$\nu_{t-1} = T_t \nu_t. \tag{25}$$

За допомогою рівнянь (24),(25) та граничної умови $\nu_n = r$ послідовно обчислюємо $\nu_n, \nu_{n-1}, \ldots, \nu_0$.

Далі для кожного $t = 1, 2, \ldots, n$ вибираємо селектор ψ_t багатозначного відображення $A(x) : X_{t-1} \to A_t$ такий, що

$$u_t(\psi_t) = \nu_{t-1}. \tag{26}$$

За **наслідком 3** проста обривна стратегія $\varphi = \psi_1 \psi_2 \ldots \psi_n$ є рівномірно оптимальною для моделі $Z^* = Z_0^*$. Рівняння (26) можна переписати у вигляді

$$T_{\psi_t} \nu_t = \nu_{t-1},$$

де оператор T_{ψ_t} переводить функції на X_t в функції на X_{t-1} за формулою

$$T_{\psi_t} f(x) = q_t[\psi_t(x)] + \sum_{X_t} p(y|\psi_t(x)) f(y) + p_t(y^*|a) c(y^*). \tag{27}$$

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

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

Доведення. Випливає з фундаментального рівняння (10), формул (11), (27) та методу математичної індукції. □

Формула (28) свідчить про таке: результат, який дає обривна стратегія $\psi_1 \psi_2 \ldots \psi_k \pi$ не зміниться, якщо перервати керування в момент часу k, взявши за фінальну плату оцінку обривної стратегії π.

8. Марковська властивість. Нехай $0 < k < n$. Припустимо, що на відрізку $[0, k]$ ми користуємося обривною стратегією ρ, а на відрізку $[k, n]$ – обривною стратегією π(тобто стратегія в похідній моделі порядку k). За аналогією до *означення 14 (добутку стратегій)* природно сказати, що використовується стратегія $\rho\pi$.

Твердження 6. *Нехай L_0 – простір шляхів на відрізку $[0, n]$, L_k – простір шляхів на відрізку $[k, n]$ і нехай $P_x^{*\rho\pi}$ – розподіл ймовірностей, що відповідає початковому стану x і обривній стратегії $\rho\pi$, і аналогічно $P_y^{*\pi}$ – розподіл ймовірностей на L_k. Тоді для кожного $\xi = \xi(x_k a_{k+1} \ldots x_n)$ з L_k правильна формула*

$$E_x^{*\rho\pi} \xi = E_x^{*\rho} [E_{x_k}^{*\pi} \xi]. \tag{29}$$

Доведення. Для кожного $l = y_0 b_1 \dots b_k y_k b_{k+1} \dots y_n$ згідно з формулою (2)

$$P_x^{*\rho\pi}(y_0 b_1 \dots y_n) = P^{*\rho}(cy_k) P_{y_k}^{*\pi}(y_k d), \tag{30}$$

де $c = y_0 b_1 \dots b_k$, $d = b_{k+1} \dots y_n$. Будь-яку функцію ξ в просторі L_k можна трактувати як функцію в просторі L_0, яка не залежить від $x_0 a_1, \dots, a_k$. Тому домножимо праву та ліву частину рівності (30) на $\xi(y_k d)$ та підсумуємо по всіх шляхах

$$E_x^{*\rho\pi}\xi = \sum_{cy_k} P_x^{*\rho}(cy_k) \sum_d P_{y_k}^{*\pi}(y_k d)\xi(y_k d). \tag{31}$$

Але $P_{y_k}^{*\pi}(yd) = 0$ при $y \neq y_k$, тому

$$\sum_d P_{y_k}^{*\pi}(y_k d)\xi(y_k d) = \sum_{yd} P_{y_k}^{*\pi}(yd)\xi(yd) = F(y_k). \tag{32}$$

Підставляємо (32) в (31), врахувавши $\sum\limits_{cy_k} P_x^{*\rho}(cy_k)F(y_k) = E_x^{*\rho}F(x_k)$, отримуємо формулу (29). □

Наслідок 5 (Марковська властивість)**.** *Нехай* $\nu(y) = P_\mu^{*\rho}\{x_k = y\}$ $(y \in X_k)$, *тоді для кожного* μ

$$E_\mu^{*\rho\pi}\xi = E_\mu^{*\rho}[E_{x_k}^{*\pi}\xi].$$

Зокрема,

$$E_\mu^{*\rho\pi}\xi(x_k a_{k+1} \dots x_n) = E_\nu^{*\pi}\xi(x_k a_{k+1} \dots x_n). \tag{33}$$

Доведення. Випливає з (29) та рівності $\sum\limits_{y \in X_k} \nu(y) P_y^{*\pi}\xi = E_\nu^{*\pi}\xi$. □

Формула (33) засвідчує, що розподіл ймовірностей для частини траєкторії на відрізку $[k, n]$ при відомому розподілі стану x_k не залежить від розподілу μ і стратегії ρ. Іншими словами, ймовірнісний прогноз "майбутнього" (ξ) якщо відоме "теперішнє" (ν) не залежить від "минулого" (μ, ρ). Це і є **марковською властивістю**.

Тепер використаємо марковську властивість для того, щоб оцінити вклади інтервалів $[0, k]$ та $[k, n]$ в оцінку обривної стратегії $\rho\pi$. Застосовуємо формулу (33) до функції $\xi = \sum\limits_{t=k+1}^{n} [q(a_t) + c(x_t^*)] + r(x_n)$ та отримуємо

$$\omega(\mu, \rho\pi) = \sum_{t=1}^{k} E_\mu^{*\rho\pi}[q(a_t) + c(x_t^*)] + \omega(\nu, \pi) = \sum_{t=1}^{k} E_\mu^{*\rho}[q(a_t) + c(x_t^*)] + \omega(\nu, \pi). \tag{34}$$

Сума в формулі (34) виражає оцінку $\omega(\mu, \rho)$ стратегії ρ при нульовій фінальній платі, тобто $\omega(\mu, \rho\pi) = \omega(\mu, \rho) + \omega(\nu, \pi)$.

Можна дати формулі (34) й іншу інтерпретацію. Згідно з (8) та $\nu(y) = P_\mu^{*\rho}\{x_k = y\}$ $(y \in X_k)$ отримуємо

$$\omega(\nu, \pi) = \sum_y \nu(y)\omega(y, \pi) = E_\mu^{*\rho}\omega(x_k, \pi),$$

$$\omega(\mu, \rho\pi) = E_\mu^{*\rho}[\sum_{t=1}^{k} q(a_t) + \omega(x_k, \pi)]. \tag{35}$$

Отже, оцінка обривної стратегії $\rho\pi$ дорівнює оцінці обривної стратегії ρ при фінальній платі $\omega(\cdot,\pi)$ в момент k.

9. Принцип динамічного програмування. Нехай Z^* – модель на відрізку $[0,n]$ і нехай $0 \leqslant s < t \leqslant n$. Позначимо $Z^*_{s,t}[f]$ – модель, яку одержують з Z^*, якщо звузити інтервал $[0,n]$ до $[s,t]$ і визначити в момент часу t фінальну плату f. Оцінку моделі Z^{*t}_s, яка відповідає фінальній платі f, позначимо $\nu^t_s[f]$. Відомо, що $\nu^t_s[f] = (VU)^{t-s}f = T^{t-s}f$ на X.

Звідси $\forall t \in [0,n]$ правильне рівняння

$$\nu^n_0[r] = \nu^t_0[\nu^n_t[r]] \text{ на } X_0 \ (r \text{ задана на } X_n). \tag{36}$$

Рівняння (36) рівносильне рівнянням оптимальності (24) та граничній умові. Воно виражає **принцип динамічного програмування**, згідно з яким для оптимізації керування на проміжку $[0,n]$ при фінальній платі r можна спочатку оптимізувати керування на проміжку $[t,n]$(при тій самій фінальній платі), а потім оптимізувати керування на проміжку $[0,t]$ при фінальній платі $\nu^n_t[r]$.

Зокрема, з рівняння (36) випливає таке: якщо π'' – рівномірно оптимальна обривна стратегія для Z^{*n}_t при фінальній платі r і π' – рівномірно оптимальна обривна стратегія для Z^{*t}_0 при фінальній платі $\nu^n_t[r]$, то обривна стратегія $\pi = \pi''\pi'$ має оцінку $\nu^n_0[r]$, а отже, рівномірно оптимальна для процесу Z^{*n}_0(при фінальній платі r).

1. *Дынкин Е.Б.* Управляемые марковские процессы и их приложения / *Дынкин Е.Б., Юшкевич А.А.* – М., 1975.
2. *Губенко Л.Г.* Об управляемых марковских процессах с дискретным временем / *Губенко Л.Г., Штатланд Э.С.* // Теор. вер. и мат. стат. – 1972. – №7. – С. 51-64.
3. *Derman C.* Finite state Markovian decision process / *Derman C.* – N. Y. – London.
4. *Bellman R.* A Markovian decision process / *Bellman R.* // J. Math. Mech. – 1964. – №6. – P. 679-684.

KILLED MARKOV DECISION PROCESSES ON FINITE TIME INTERVAL FOR FINITE MODELS

Nestor PAROLYA, Yaroslav YELEJKO

Ivan Franko National University of L'viv,
79000 L'viv, Universytets'ka Str., 1
e-mail: mstat@franko.lviv.ua

In this article we consider killed Markov decision processes for finite models on finite time interval. Existence of a uniform optimal strategy is proved. We showed correctness of the fundamental equation. Optimal control problem is reduced to a similar problem for derived model. We receive optimality equation and method for simple optimal strategies constructing. We show correctness

of the Markovian property. Additionally dynamic programming principle is considered.

Key words: Markov decision process, optimal strategy, optimal decision, fundamental equation.

ОБРЫВНЫЕ УПРАВЛЯЕМЫЕ МАРКОВСКИЕ ПРОЦЕССЫ НА КОНЕЧНОМ ИНТЕРВАЛЕ ВРЕМЕНИ ДЛЯ КОНЕЧНЫХ МОДЕЛЕЙ

Нестор ПАРОЛЯ, Ярослав ЕЛЕЙКО

Львовский национальный университет имени Ивана Франко,
79000 Львов, ул. Университетская, 1
e-mail: mstat@franko.lviv.ua

Рассмотрено обрывные управляемые марковские процессы для конечных моделей на конечном интервале времени. Доказано существование равномерной оптимальной стратегии. Обосновано справедливость фундаментального уравнения. Приведено задачу оптимального управления к аналогичной задачи для производной модели. Получено уравнения оптимальности и метод построения простых равномерных оптимальных стратегий. Показано справедливость марковского свойства. Рассмотрено принцип динамического программирования.

Ключевые слова: управляемый марковский процесс, оптимальная стратегия, оптимальное управление, фундаментальное уравнение.

Стаття надійшла до редколегії 26.01.2010

Прийнята до друку 22.12.2010