УДК 519.21

ОБ ОДНОЙ МОДЕЛИ УПРАВЛЯЕМЫХ ПОЛУМАРКОВСКИХ ПРОЦЕССОВ

***П.С.Кнопов, **В.А.Пепеляев**

Институт кибернетики им.В.И.Глушкова НАН Украины,
*отдел математических методов исследования операций,
**отдел методов системного моделирования,
ул. Академика В.М.Глушкова, Киев-187, 03187,
e-mail: **knopov1@yahoo.com, **pepelaev@yahoo.com*

Abstract

The problem of optimal strategy founding of the repair organization for the system that is described bi the semi-Markov process is investigated. The statements about the optimal strategy existence are proved, the optimal strategy structure is found.

Введение

В данной работе рассматриваются вопросы оптимизации ремонтных работ с помощью теории управляемых полумарковских процессов. Для исследования этих моделей нам понадобятся некоторые утверждения о существовании оптимального управления полумарковскими системами в компактных пространствах состояний и управлений. Для марковских моделей такие утверждения приведены в [1], и мы будем обращаться к ним по мере необходимости. Наше дальнейшее изложение ограничивается рассмотрением только полумарковских процессов. Важной особенностью этих процессов, является то, что время перехода из одного состояния в другое есть некоторая случайная величина. Этот факт делает возможным исследование довольно широкого класса приложений. Впервые в рассмотрение полумарковские процессы были введены и подробно изучены П.Леви [2] и В.Смитом [3].

Управляемые полумарковские процессы с конечными множествами состояний и управлений были изучены С.Дерманом [4] и В.Джевелом [5]. Случай самых общих множеств состояний и управлений рассматривались С.Россом [6], Л.Г.Губенко и Э.С.Штатландом [7], Х.Гуо и О.Хернандез-Лерма [8].

1. Необходимые сведения из теории управляемых случайных процессов

Напомним некоторые факты из теории управляемых полумарковских скачкообразных процессов, которые будем использовать в дальнейшем.

Рассмотрим систему со случайными внешними воздействиями, эволюционирующей в непрерывном времени как случайный скачкообразный процесс, который в моменты скачкообразного изменения процесса является объектом управления для лица, принимающего решения (ЛПР). Пространство состояний (фазовое пространство) стохастического процесса $X = (X_t : t \in \mathbb{R}_+)$, описывающего эволюцию системы во времени обозначим $\mathbf{X}$, а пространство решений - $\mathbf{A}$. Будем считать, что как $\mathbf{X}$, так и $\mathbf{A}$ являются сепарабельными, полными метрическими пространствами с заданными на них борелевскими σ - алгебрами Ξ и $\mathcal{A}$ соответственно.

Если после n-го скачка в момент S_n, $n \in \mathbb{N}$, система находится в состоянии $x \in X$, то ЛПР может принимать решение $D_n = a, \quad a \in A_x \in \mathcal{A}$, где A_x есть допустимое множество решений в состоянии x.

Обозначим

$$A : \mathbf{X} \to \mathcal{A}, \quad x \to A_x$$

отображение, связывающее каждое состояние с множеством возможных в этом состоянии решений. При этом очевидно [9], что $\Delta = (x, a), x \in X, a \in A_x$ есть борелевское измеримое подмножество в пространстве $\mathbf{X} \times \mathbf{A}$.

Случайная эволюция системы описывается множеством вероятностей переходов из одного состояния в следующее, зависящих от времени между смежными переходами.

Последовательность скачков процесса имеет переходные вероятности

$$P\{B/x_n, a_n\} = P\{X_{n+1} \in B/X_0 = x_0, D_0 = a_0, \tau_0 \leqslant t_0, \dots, X_n = x_n, D_n = a_n\}$$

где $B \in \Xi, (x_k, a_k) \in \Delta \in (\Xi \times \mathcal{A})$, а $x_k := X_{S_{k+}}$ есть состояние системы в момент S_k, a_k - выбранное решение в момент S_k и $\tau_k = S_{k+1} - S_k$

Распределение промежутков времени задается следующим образом. Если система находится в состоянии $x \in \mathbf{X}$, принимается решение $a \in \mathbf{A}$, и следующим состоянием является $y \in \mathbf{X}$ (выбрано в соответствии с $P\{\cdot/x, a\}$), тогда случайный промежуток времени τ в состоянии x имеет функцию распределения $\Phi(\cdot/x, a, y) = P(\tau \leqslant \cdot/x, a, y)$. Предполагается, что функции $P(\cdot/x, a)$ и $\Phi(\cdot/x, a, y)$ измеримы на Δ и $\Delta \times X$ соответственно.

Структура стоимости в управляемой системе следующая. Обозначим через $r(s/x, a)$ ожидаемый доход за один промежуток времени, если система находится в состоянии x в начале периода, принято решение $a \in A_x$, и промежуток времени в состоянии x есть $s \in R_+$. Предполагается, что $r(s/x, a)$ является ограниченной измеримой функцией на $[0, \infty] \times \Delta$.

Допустимой стратегией управляемой системы является последовательность переходов $\delta = \{\pi_0, \pi_1, \dots, \pi_n\}$ таких, что вероятностная мера $\pi_n(\cdot/x_0, a_0, \tau_0, \dots, x_n)$ на $(A \times \mathcal{A})$ сконцентрирована на A_{x_n} (подробнее см. [1]).

Стратегия δ называется стационарной марковской, если $\pi_n(\cdot/x_0, a_0, \tau_0, \dots, x_n) = \pi_n(\cdot/x_n), \quad n = 0, 1, 2, \dots$ Стационарная марковская стратегия является детерминированной (сокращенно: стационарной детерминированой), если мера $\pi_n(\cdot/x_n)$ точечная для любого $x \in X$.

Обозначим класс всех допустимых стратегий через $\mathfrak{R}$, а класс стационарных марковских детерминированных стратегий через $\mathfrak{R}_1$.

Рассмотрим среднюю стоимость на бесконечности для оценивания производительности системы при стратегии, описанной выше. Она определяется выражением:

$$\varphi_\delta(x) = \underline{\lim}_{n\to\infty} \frac{E_x^\delta \sum_{k=0}^n r(\tau_k/x_k, a_k)}{E_x^\delta \sum_{k=0}^n \tau_k}, \tag{1}$$

где E_x^δ - условное ожидание для системы, находящейся в состоянии x в начальный момент времени 0 и использующей стратегию δ.

Стратегия δ^* называется оптимальной, если

$$\varphi_{\delta^*}(x) = \sup_{\delta\in\Re} \varphi_\delta(x), \qquad x \in X$$

Обозначим ожидаемый промежуток времени в состоянии x при решении a

$$\tau(x,a) = \int_X \int_0^\infty t d\Phi(t/x,a,y) P(dy/x,a),$$

и ожидаемую стоимость в состоянии x при решении a

$$r(x,a) = \int_X \int_0^\infty r(t/x,a) d\Phi(t/x,a,y) P(dy/x,a)$$

Подразумевается, что интегралы существуют, и $\tau(x,a)$ и $r(x,a)$ конечны для произвольных $(x,a) \in \Delta, |r(x,a)| \leqslant K < \infty$.

Замечание 1. Функция (1) зависит только от переходных вероятностей последовательности скачков $P\{\cdot/x,a\}$ и ожиданий $\tau(x,a)$ и $r(x,a)$. Поэтому (1) нечувствительна к средним отклонениям в распределениях процесса до тех пор, пока эти вероятности переходов и средние ожидания остаются неизменными.

Замечание 2. В наших моделях мы рассматриваем системы, для которых имеет место следующее

$$\Phi(t/x,a,y) = \begin{cases} 1, & t \geqslant \tau(x,a), \\ 0, & t < \tau(x,a), \end{cases}$$

$$r(t/x,a) = \begin{cases} r(x,a), & t \geqslant \tau(x,a), \\ 0, & t < \tau(x,a). \end{cases}$$

Далее нам понадобится следующее утверждение [7].

Теорема 1. *Пусть пространство состояний* $\mathbf{X}$ *пространство решений* $\mathbf{A}$ *управляемой системы, описанной выше, являются компактными, и пусть выполняются условия:*

1) $0 < l \leqslant \tau(x,a) \leqslant L < \infty, \quad (x,a) \in \Delta$

2) *существует неотрицательная мера* μ *на* (X, Ξ) *такая, что* $\mu(X) > 0$ *и* $\mu(B) \leqslant P\{B/x, a\}, (x, a) \in \Delta, B \in \Xi$*;*
3) *отображение* A*, которое преобразует произвольную* $x \in \mathbf{X}$ *в непустое множество* $A_x \in \mathcal{A}$ *является полунепрерывным сверху;*
4) *переходные вероятности* $P\{\cdot/x, a\}$ *слабо непрерывны на* (x, a)*;*
5) *функция* $r(x, a)$ *полунепрерывна сверху, и функция* $\tau(x, a)$ *непрерывна на* $(x, a) \in \Delta$*.*

Тогда в классе $\Re_1$ *существует оптимальная стратегия, которая достигает максимальной стоимости*

$$W = \frac{1}{L} \int V(x) \mu(dx),$$

где $V(x)$ *- решение уравнения оптимальности*

$$V(x) = \sup_{a \in A_x} \left\{ r(x, a) + \int V(y) P'(dy/x, a) \right\},$$

где

$$P'(B/x, a) = P(B/x, a) - \frac{1}{L} \mu(B) \tau(x, a)$$

Величина $V(x)$ единственным образом определена на $\mathcal{B}(\mathbf{X})$, банаховом пространстве ограниченных борелевских измеримых функций на $\mathbf{X}$ с нормой $\|u\| = \sup_{x \in X} |u(x)|$ и может быть получена методом последовательных приближений.

Рассмотрим еще один критерий оптимальности системы. Для некоторого множителя $\alpha < 1$, начального состояния $x \in X$ и стратегии δ определим дисконтный общий выигрыш на бесконечности

$$\psi_\delta(x, a) = E_x^\delta \sum_{n=0}^{\infty} r(x_n, a_n) exp \left\{ -\alpha \sum_{k=0}^{\infty} \tau(x_k, a_k) \right\} \tag{2}$$

Теорема 2. ([7]) *Пусть* $\mathbf{X}$ *компактно, и выполнены следующие условия:*

1) $\tau(x, a) \geqslant \ell > 0, (x, a) \in \Delta$*;*
2) A *компактно или конечно;*
3) *функции* $r(x, a)$ *и* $\tau(x, a)$ *непрерывны на* Δ*;*
4) *переходная вероятность* $P(\cdot/x, a)$ *слабо непрерывна на* Δ*.*

Тогда в классе $\mathfrak{R}_1$ *существует оптимальная стратегия* δ_α *относительно дисконтного критерия стоимости (2), при которой достигается оптимальный доход* $\psi_{\delta_\alpha}(x, \alpha) = \psi_\alpha(x)$*, который является единственным решением уравнения*

$$\psi_\alpha(x) = \max_{a \in A} \left\{ e^{-\alpha \tau(x, a)} \left[r(x, a) + \int \psi_\alpha(y) P(dy/x, a) \right] \right\}$$

в пространстве $\mathcal{B}(\mathbf{X})$*.*

Заметим, что метод доказательства теоремы 1 состоит в получении уравнения оптимальности для $\varphi-$критерия и аппроксимированного $\psi_\alpha-$критерия переходом к пределу $\alpha \to 0$.

2. Нахождение оптимальной стратегии ремонтных работ

Пусть эволюция системы описывается равномерно монотонным невозрастающим марковским процессом $\{\xi(t), t \geqslant 0\}$ со значениями в $[0, \infty]$ и переходными вероятностями $P(x, t, B), x \in [0, \infty], t \geqslant 0, B \in \mathbf{B}_+$, борелевским множеством на $[0, \infty]$. Состояние $\{0\}$ соответствует исправному состоянию системы, а $x > 0$ характеризует некоторый уровень неполадки.

Текущее состояние системы может быть определено непосредственной проверкой со стоимостью затрат $r_1 > 0$. Стоимость производимой продукции зависит от состояния системы. Будем предполагать, что стоимость продукции в состоянии x равна $r(x) \geqslant 0$. Функция $r(x)$ монотонная, невозрастающая и ограниченная.

В зависимости от состояния системы после каждой проверки необходимо принимать решение: или мы ничего не предпринимаем и осуществляем следующий контроль через время T (это действие обозначается a_T), или производим полную замену, которая продолжается $m \geqslant 0$ единиц времени, а следующая проверка осуществляется через время T после возобновления работы системы в состоянии $\{0\}$ (это действие обозначим (a_0, a_T)).

Предполагается, что длина интервала времени T между проверками принадлежит некоторому множеству $Z \subseteq [T_1, T_2], 0 < T_1 < T_2$, где Z или конечно или весь интервал $[T_1, T_2]$. Предполагается также, что ремонт длится m единиц времени и ремонт в единицу времени стоит $r_2 > 0$. Ремонт включает собственную стоимость ремонта, а также потери из-за неисправности системы. Предполагается, что имеет место $\lim_{x\to\infty} r(x) > r_2$.

Введем величину $c = \min\{x : x \geqslant 0, r(x) \geqslant r_2\}$. Состояния $x \geqslant c$ будем называть состояниями неисправности системы. Если в момент проверки состояния процесса $x \geqslant c$, то всегда принимается решение (a_0, a_T).

Эти предположения сводят нашу модель к простой управляемой полумарковской модели с критерием (1), пространством состояний $\mathbf{X} = [0, c]$, пространством управлений $\mathbf{A} = \{a_T, (a_0, a_T)\}$, $A_x = \begin{cases} \mathbf{A}, & 0 \leqslant x < c, \\ \{(a_0, a_T)\}, & x = c, \end{cases}$ и следующими вероятностями переходов

$$P(B/x, a_T) = P(x, T, B),\ x \in [0, c),\ B \in [0, c) \cap \mathbf{B}_+,$$

$$P\{\{c\}/x, a_T\} = P(x, T, [c, \infty)),\ x \in [0, c),$$

$$P\{B/x, (a_0, a_T)\} = P(0, T, B),\ x \in [0, c],\ B \in [0, c) \cap \mathbf{B}_+,$$

$$P\{\{c\}/x, (a_0, a_T)\} = P(0, T, [c, \infty)),\ x \in [0, c).$$

Заметим, что вероятность P в различных частях уравнений имеет различный смысл. Однако это не должно привести к затруднениям.

Среднее время продолжительности процесса определяется следующим образом:

$$\tau(x, a_T) = T,\ x \in [0, c),$$

$$\tau(x, (a_0, a_T)) = m + T,\ x \in [0, c),$$

а средняя полезная стоимость задается посредством

$$r(x, a_T) = -\left[r_1 + E_x \int\limits_0^T r(\xi(t))dt \right],\ x \in [0, c),$$

$$r(x, (a_0, a_T)) = -\left[r_1 + E_0 \int\limits_0^T r(\xi(t))dt + r_2 m \right],\ x \in [0, c),$$

где E_x условное ожидание, соответствующее мере процесса $\xi = (\xi(t) : t \in \mathbb{R}_+)$, при условии $\xi(0) = x$. Предположим далее, что переходные вероятности $P(x, T, B)$ слабо непрерывны на (x, t), $P(0, T_1, [c, \infty)) = \gamma > 0$.

Теорема 3. *Пусть для произвольной борелевской функции и на* $[0, \infty)$ *функция* $E_x(u(\xi(t)))$ *непрерывна по* x *и по* t. *Тогда для модели, описанной выше, существует оптимальная стратегия* $\varphi_\delta(\mathrm{x})$ *в классе* $\Re_1$, *для которой достигается максимальное значение стоимости* $W = \frac{1}{T+m} \int V(x)\mu(dx)$, *где* $\mu(\cdot)$ *— мера, сосредоточенная в точке 0 с массой* γ, *и* $V(x)$ *удовлетворяет уравнению оптимальности*

$$V(x) = \max[-r_1 - E_x \int\limits_0^T r(\xi(t))\,dt + \int V(y)P'(dy/x, a_T),$$

$$-r_1 - E_0 \int\limits_0^T r(\xi(t))\,dt + \int V(y)P'(dy/x, (a_0, a_T)) - r_2 m] = FV(x) \qquad (3)$$

или

$$V(x) = \max\Big[-r_1 - E_x \int\limits_0^T r(\xi(t))\,dt + \int\limits_x^c V(y)P(x, T, dy) +$$

$$+ V(c)P(x, T, [c, \infty)) - \frac{\gamma T V(0)}{T + m},$$

$$-r_1 - E_0 \int\limits_0^T r(\xi(t))\,dt + \int\limits_0^c V(y)P(0, T, dy) +$$

$$+V(c)P(0,T,[c,\infty)) - \frac{\gamma(T+m)V(0)}{T+m} - r_2 m\Big]. \tag{4}$$

Доказательство. Для отображения A из множества состояний $\mathbf{X}$ в множество управлений $\mathbf{A}$ имеем: если $x_n \to x$, $\lim_{n\to\infty} a_n = a$, $x_n \in \mathbf{X}$, $a_n \in A_{x_n}$ тогда $a \in A_x$, поэтому A полунепрерывно сверху. Таким образом условие 3 теоремы 1 выполнено. Другие условия теоремы 1 могут быть проверены непосредственно. Поэтому в классе детерминированных стратегий существует стратегия, при которой максимальный доход W достигается, и W может быть получена методом последовательных приближений.

Наша дальнейшая цель - определить структуру оптимальной стратегии. Обозначим для промежутка времени между проверками T

$$u_T(x) = u_{T_1}(x) - u_{T_2}(x),$$

где

$$u_{T_1} = u_1(x) = \Big[-r_1 - E_x \int\limits_0^T r(\xi(t))\,dt + \int\limits_x^c V(y)P(x,T,dy) +$$

$$+V(c)P(x,T,[c,\infty)) - \frac{\gamma T V(0)}{T+m}\Big],$$

$$u_{T_2} = u_2(x) = \Big[-r_1 - E_0 \int\limits_0^T r(\xi(t))\,dt + \int\limits_0^c V(y)P(0,T,dy) +$$

$$+V(c)P(0,T,[c,\infty)) - \frac{\gamma(T+m)V(0)}{T+m} - r_2 m\Big].$$

В силу наших предположений получаем, что $u_T(0) = \frac{\gamma m V(0)}{T+m} + r_2 m > 0$. В состоянии $x \geqslant c$ всегда выбирается решение (a_0, a_T), поэтому $u_T(c) < 0$. □

Отсюда непосредственно вытекает следующее утверждение.

Теорема 4. *Пусть для произвольной монотонно невозрастающей ограниченной функции $u(x)$, заданной на $[0,\infty)$, функция $E_x(u(\xi(t)))$ монотонно невозрастающая по x для любого $t \geqslant 0$. Тогда оптимальная стратегия $\delta^* \subset \Re_1$ имеет вид*

$$\delta^* = \begin{cases} a_T, x < x^*, \\ (a_0, a_T), x \geqslant x^*, \end{cases} \tag{5}$$

где $x^* \in [0,c]$.

Доказательство. Сначала докажем, что $V(x)$ монотонно невозрастающая. В условиях теоремы функция $-E_x \int\limits_0^T r(\xi(t))\,dt$ монотонно невозрастающая по x. Запишем уравнение оптимальности в виде

$$V(x) = FV(x) = \max\Bigg[-r_1 - E_x\int\limits_0^T r(\xi(t))\,dt + E_xV(\xi(T)) - \frac{\gamma TV(0)}{T+m},$$

$$-r_1 - E_0\int\limits_0^T r(\xi(t))\,dt + E_0V(\xi(T)) - \frac{\gamma(T+m)V(0)}{T+m}\Bigg].$$

Для нахождения $V(x)$ мы имеем $V(x) = \lim\limits_{n\to\infty} F^nV^0(x) = \lim V^{(n)}(x)$, где $V^0(x)$ — произвольная функция, а $V^n(x)$ определяется следующим образом

$$V^n(x) = \max\Bigg[-r_1 - E_x\int\limits_0^T r(\xi(t))\,dt + E_xV^{(n-1)}(\xi(T)) - \frac{\gamma TV^{(n-1)}(0)}{T+m},$$

$$-r_1 - E_0\int\limits_0^T r(\xi(t))\,dt + E_{x0}V^{(n-1)}(\xi(T)) - \frac{\gamma(T+m)V^{(n-1)}(0)}{T+m}\Bigg].$$

Используя в качестве $V^0(x)$ произвольную монотонно невозрастающую функцию, получим, что $V(x)$ также невозрастающая. Поэтому $u_T(x)$ также монотонно невозрастающая на $x \in [0, c]$. Более того, $u(0) > 0$ и $u(c) < 0$.

Следовательно, существует единственная точка $x^* \in [0, c]$ такая, что оптимальная стратегия $\delta^* \subset \Re_1$ имеет вид (5). $\square$

Замечание 3. Случай ψ_α-критерия исследуется аналогичным образом. Условия существования оптимальной стратегии в классе $\Re_1$, аналогичны условиями теоремы 3.

Заключение

Таким образом, *в работе получены условия существования оптимальной стратегии в классе марковских стационарных детерминированных управлений, и найден вид этой стратегии, имеющей простую двухуровневую структуру.*

Список литературы

1. *Губенко Л.Г., Штатланд Э.С.* Об управляемых марковских процессах // Теория вероятности и математ. статистика, Изд-во КГУ. - 1972, вып.7. - С.51-64.
2. *P.Levy* Processus semi-markoviens// Proc. Int. Cong. Math. (Amsterdam). - 1954, Vol.3. - P.416-426.
3. *W.L.Smith* Regenerative stochastic process// Proc. Roy. Soc. A. - 1955, Vol.232. - P.6-31.
4. *C.Derman* On sequential decisions and Markov chain // Manag. Sci. - 1962, Vol.9. - P.16-24.
5. *W.S.Jewell* Markov-Renewal programming, I, II// Operation Research. - 1963, Vol.11, no 6. - P.938-971.
6. *S.Ross* Average cost Semi-Markow decision processes//Journal of Applied probability. - 1970, Vol.7, no 3.

7. *Губенко Л.Г., Штатланд Э.С.* Об управляемых полумарковских процессах//Кибернетика. - 1972, №2. - С.26-29.
8. *X.Guo, O.Hernandez-Lerma* Continuous-time controlled Markov chains// Annals of Applied Probability. - 2003, 13. - P.363-388.
9. *R.T.Rocafellar* Measurable Dependence of Convex Sets and Functions on Parameters//Journal of Mathematical Analysis and Applications. - 1969, 28, 1. - P.4-25.