Управляемые марковские цепи в условиях риска

Ибрагимов А.А.
Кафедра геометрии и прикладной математики,
Национальный университет Узбекистана,
100174, ВУЗ городок, г.Ташкент,
Узбекистан
infinity_x@mail.ru

Аннотация

Рассматриваются управляемые марковские цепи с несколькими эргодическими классами и классом невозвратных состояний, которые могут меняться в зависимости от стратегии. Задача максимизации предельного среднего дохода изучается с позиции теории статистических решений и полезности, и предлагаются критерии риска, позволяющие найти предпочтительные решения.

1. Пусть имеется управляемая марковская цепь с конечным множеством состояний $S = 1, 2, \dots, N$ и конечными множествами решений $F_i = 1, 2, \dots, k_i$ $(i \in S)$. Элемент f из $F = F_1 \times F_2 \dots \times F_N$ — решающая функция; $\varphi = (f_1, f_2, \dots, f_n, \dots,)$ — стратегия; $f^\infty = (f, f, \dots, f, \dots,)$ — стационарная стратегия. Для каждой $f \in F$ заданы $(N \times N)$-матрица переходных вероятностей $P(f) = \|p_{ij}^k\|(i, j \in S)$ и $(N \times 1)$-мерный вектор доходов $r(f) = [r_i^k](i \in S)$, где $k = f(i) \in F_i$.

Цель управления состоит в максимизации $(N \times 1)$-мерного вектора предельных средних доходов (средних доходов за один шаг)

$$G(\varphi) = \lim_{T \to \infty} \inf \frac{1}{T} \sum_{n=0}^{T-1} P_n(\varphi) r(f_{n+1}) \tag{1}$$

в множестве стратегий φ, где $P_n(f) = P(f_1)P(f_2)\dots P(f_n)$, $P_0(\varphi) = I$ (I — единичная $(N \times N)$-матрица). Здесь i-я координата $g_i(\varphi)$ вектора $G(\varphi)$ отвечает начальному состоянию процесса $i \in S$.

Доказано, что данная задача оптимального управления имеет решение в классе стационарных стратегий f^∞. Разработаны алгоритмы нахождения стационарной оптимальной стратегии. Результаты изложены в книге Майна и Осаки (1977).

2. При фиксированном $f \in F$ множество состояний S цепи Маркова, задаваемой матрицей $P(f)$, разбивается на несколько эргодических множеств $E_1, E_2, \dots, E_m$, и множество невозвратных состояний E_0. Последнее характеризуется тем, что оказавшись в этом множестве, цепь с вероятностью 1 за конечное число шагов покидает его, поглощаясь в одном из эргодических множеств.

Пусть для любого ν $(\nu = \overline{1, m})$ цепь $(E_\nu, P_\nu(f))$, где $P_\nu(f)$ — подматрица матрицы $P(f)$, соответствующая эргодическому классу E_ν, является регулярной цепью Маркова. Тогда существует предел $\lim_{n\to\infty} P_\nu^n(f) = P_\nu^*(f)$, где $P_\nu^*(f)$ — матрица с одинаковыми строками, образованными из предельных вероятностей состояний $\pi_j(\nu, f) > 0, j \in E_\nu$. Отсюда приходим к выражению предельного среднего дохода

$$g_i(f) = \begin{cases} g(\nu, f) = \sum\limits_{j \in E_\nu} \pi_j(\nu, f) r_j^k, & \text{если } i \in E_\nu, \\ \sum\limits_{\nu=1}^{m} b_{i\nu}(f) g(\nu, f), & \text{если } i \in E_0, \end{cases} \tag{2}$$

где $b_{i\nu}(f)$ — вероятность поглощения цепи, исходящей из невозвратного состояния E_0, в ν-м эргодическом множестве E_ν.

Матрица вероятностей поглощения $B(f) = \|b_{i\nu}(f)\|(i \in E_0,\ \ \nu = \overline{1, m})$ определяется формулой $B(f) = [I - Q(f)]^{-1} R(f)$, где квадратная матрица $Q(f) = \|p_{ij}^k\|(i, j \in E_0)$, описывает поведение цепи до выхода из множества невозвратных состояний, а матрица $R(f) = \|\widetilde{p}_{i\nu}^{\,k}\| (i \in E_0, \;\; \nu = \overline{1,m})$, где $\widetilde{p}_{i\nu}^{\,k} = \sum_{j \in E_\nu} p_{ij}^k$, отвечает переходам из невозвратных в эргодические множества.

3. П р и м е р 1. Рассмотрим управляемую цепь с матрицей и доходами

$$P = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline p & q & 1-p-q \end{array} \right), \quad r = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}, \quad r_1 < r_2.$$

Вероятности поглощения равны: $b_{31} = 1 - \beta, \; b_{32} = \beta$, где $\beta = p/(p+q)$.

Цепь характеризуется тремя состояниями, два из которых поглощающие, а третье, в зависимости от принятого решения, может быть либо невозвратным ($p > 0, \; q > 0$), либо поглощающим ($p = 0, \; q = 0$). Согласно (2) предельный средний доход цепи, вышедшей из состояния 3, равен

$$g_3 = \begin{cases} \beta r_2 + (1-\beta), & \text{если } p > 0, \; q > 0, \\ r_3, & \text{если } p = 0, \; q = 0. \end{cases}$$

Если $r_1 > r_3$, то оптимально первое решение; если $r_2 < r_3$, то оптимально второе решение; если $r_1 < r_3 < r_2$, то возникает задача принятия решений в условиях риска. Положим: $r_1 = 0, \; r_2 = 10^6, \; r_3 = 50$; решение a: $p = 0, \; q = 0$; решение b: $p = 1 - 10^{-4}, \; q = 10^{-4}$. Тогда $\beta = 1 - 10^{-4}, \; g_3(a) = 50, \; g_3(b) = 100$.

По критерию (1) решение b оптимально. Однако, при нем цепь наверняка (с вероятностью $1 - 10^{-4}$) окажется в первом состоянии, где доход за один шаг равен нулю. Здесь "ожидаемый"доход за один шаг 100 у.е. в действительности не существует. Принять решение b в третьем состоянии не разумно, в нем предпочтительнее решение a, при котором доход 50 у.е. за один шаг гарантирован 100%.

4. Принятие решения b может быть истолковано как проведение эксперимента, исходом которого является попадание цепи к одному из поглощающих состояний 1 или 2. Такое действие равносильно отказу от гарантированного дохода за один шаг 50 у.е. Эту сумму можно рассмотреть как плату за право экспериментировать. С этой точки зрения, пример 1 представляет собой статистическую задачу решения с входной платой, изученной Ибрагимовым (1997) с позиции теории статистических решений и полезности. Изложение теории можно найти в книгах Де Грота (1974) и Фишберна (1978).

Рассмотрим множество точек (x, y) на плоскости Oxy, для которых выполняются неравенства $r_1 \le x \le r_2, \;\; 0 \le y \le 1$; его обозначим $\mathcal{M}$. Здесь x — входная плата за проведение эксперимента, $r_i \;\; (i = 1, 2)$ — доход, получаемый при i-м исходе эксперимента, $y = \beta$ — вероятность появления второго исхода эксперимента.

Функции полезности $U(x)$ с областью определения $[r_1, r_2]$ и областью значений [0,1] удовлетворяет условиям $U(r_1) = 0, \; U(r_2) = 1$. График функции $U(x)$, как кривая $\mathcal{L}$ целиком лежит в прямоугольнике $\mathcal{M}$ и делит его на две части. Кривая $\mathcal{L}$ представляет собой границу риска, т. е. состоит из таких точек $(x^0, y^0) \in \mathcal{M}$, что отказ от проведения эксперимента, и тем самым сохранение входной платы x^0, выгоден настолько, насколько можно идти на риск — проводить эксперимент и увеличить имеющийся капитал x^0 до r_2 с вероятностью $\beta = y$ или снизить его до r_1 с вероятностью $1 - \beta$.

Граница риска $\mathcal{L}$ характеризуется тем, что все точки $(x^+, y^+) \in \mathcal{M}$, одобряющие риск лежат над $\mathcal{L}$, а все точки $(x^-, y^-) \in \mathcal{M}$, опровергающие риск лежат под $\mathcal{L}$. Так что, $y^+ > U(x^+), \; y^- < U(x^-), \; y^0 = U(x^0)$.

Роль функции полезности могут выполнить следующие функции:

$$U_1(x) = (x - r_1)/(r_2 - r_2),$$

$$U_2(x) = \begin{cases} \sqrt{z_x/2}, & \text{если } r_1 \le x \le \overline{r}, \\ 1 - \sqrt{z_x/2}, & \text{если } \overline{r} \le x \le r_2, \end{cases} \quad U_3(x) = \begin{cases} \sqrt{z_x(1-z_x)}, & \text{если } r_1 \le x \le \overline{r}, \\ 1 - \sqrt{z_x(1-z_x)}, & \text{если } \overline{r} \le x \le r_2, \end{cases}$$

где $z_x = U_1(x),\ \overline{r} = (r_1 + r_2)/2$.

График $\mathcal{L}_1$ функции $U_1(x)$ — диагональ прямоугольника $\mathcal{M}$; график $\mathcal{L}_2$ функции $U_2(x)$ — части двух парабол с вершинами $(r_1, 0)$ и $(r_2, 1)$, соответственно, пересекающихся в точке $(\overline{r}, 1/2)$; график $\mathcal{L}_3$ функции $U_3(x)$ — части двух эллипсов с вершинами $(r_1, 0), (\overline{r}, 1/2)$ и $(\overline{r}, 1/2), (r_2, 1)$, соответственно.

5. В примере 1 критерий оптимальности $G(f^\infty)$ по существу представляет собой функцию полезности $U_1(x)$. Действительно, из равенства $x = U_1(x)r_2 + [1 - U_1(x)]r_1$ следует отношение

$$x \gtreqless \beta r_2 + (1-\beta) r_1, \tag{3}$$

эквивалентное отношению $U_1(x) \gtreqless \beta$, где символ $\gtreqless$ представляет собой один из знаков $>$, $=$, $<$. Риск допускается, если $\beta > U_1(x)$. По этому критерию риска решение b оптимально, так как $10^{-4} > U_1(50) = 0,5 \cdot 10^{-4}$. Между тем, критерии риска $U_2(x)$ и $U_3(x)$ предпочтение отдадут решению a: $10^{-4} < U_2(50) = 10^{-2} < U_3(50) \approx 0,7 \cdot 10^{-2}$.

6. Теперь рассмотрим случай, когда в управляемых марковских цепях число эргодических классов больше двух.

П р и м е р 2. Рассмотрим управляемую цепь с матрицей и доходами

$$P(k) = \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline p_{51}^k & p_{52}^k & p_{53}^k & p_{54}^k & p_{51}^k \end{array}\right), \ k = \overline{1,5}, \quad r = \begin{pmatrix} 8 \\ 16 \\ 64 \\ 128 \\ 56 \end{pmatrix}.$$

Решениям из $F_5 = 1,\ \ldots,\ 5$ соответствуют следующие вероятности поглощения $B(k) = [b_{51}^k,\ b_{52}^k,\ b_{53}^k,\ b_{54}^k]$: $B(1) = [1/4,\ 1/4,\ 1/4,\ 1/4]$, $B(2) = [1/8,\ 3/8,\ 1/4,\ 1/4]$, $B(3) = [3/16,\ 3/8,\ 1/16,\ 3/8]$, $B(4) = [1/4,\ 1/16,\ 5/8,\ 1/16]$, $B(5) = [0,\ 0,\ 0,\ 0]$. Согласно (2) предельные средние доходы $g_5(k)$, $k = \overline{1,5}$ равны: $g_5(1) = 54$, $g_5(2) = 55$, $g_5(3) = 59,5$, $g_5(4) = 51$, $g_5(5) = 56$. Отсюда следует, что по критерию (1) для решений F_5 справедлива цепочка отношений предпочтения $3 \succ 5 \succ 2 \succ 1 \succ 4$.

Здесь следует обратить внимание вероятностям прибыльных исходов $\beta_5(k) = b_{53}^k + b_{54}^k$: $\beta_5(1) = 1/2$, $\beta_5(2) = 1/2$, $\beta_5(3) = 7/16$, $\beta_5(4) = 11/16$, $\beta_5(5) = 0$. По критерию (1) предпочтительным является рискованное решение $3 \in F_5$, и это несмотря на то, что при нем вероятность прибыльных исходов $\beta_5(3)$ самая низкая. В то же время решение $4 \in F_5$ с наибольшей вероятностью прибыльных исходов $\beta_5(4)$ находится в самом конце в шкале предпочтения.

7. Функции полезности $U_l(x), l = 1, 2, 3$ могут быть использованы для выбора предпочтительных решений и в случае, когда в управляемых марковских цепях число эргодических классов больше двух.

Пусть существуют решения $f'(i) \in F_i,\ i \in E_0$, при которых множество невозвратных состояний E_0 превратится в эргодический класс с предельным средним доходом $g(0, f')$, где в решающей функции $f' \in F$ решения $f'(i) \in F_i,\ i \in SE_0$ те же самые, что и в решающей функции $f \in F$. Заметим, что аналогичная ситуация возникает и в случае, когда некоторый эргодический класс E_ν при некоторых решениях $f'(i) \in F_i,\ i \in E_\nu$ превратится в множество или подмножество невозвратных состояний. Не теряя общности можно предположить, что предельные средние доходы $g(\nu, f),\ \nu = \overline{1,m}$, получаемые в эргодических классах $E_\nu,\ \nu = \overline{1,m}$, удовлетворяют условию $g(1,f) < g(2,f) < \ldots < g(m,f)$. Неравенства $g(\mu, f) < g(0,f) < g(\mu+1, f)$ определяют прибыльные $1,\ 2,\ \ldots,\ \mu$ и убыточные исходы $\mu+1,\ \mu+2,\ \ldots,\ m \quad (\mu \geq 1)$. Положим

$$d_i(f) = \sum_{\nu=1}^{\mu} \delta_{i\nu}(f) g(\nu, f), \quad D_i(f) = \sum_{\nu=\mu+1}^{m} \Delta_{i\nu}(f) g(\nu, f),$$

где $\delta_{i\nu}(f)$ $(1 \leq \nu \leq \mu)$ — вероятность появления ν-го убыточного исхода, при условии, что результатом эксперимента является убыточный исход, $\Delta_{i\nu}(f)$ $(\mu+1 \leq \nu \leq m)$ — вероятность появления ν-го прибыльного исхода, при условии, что результатом эксперимента является прибыльный исход; $i \in E_0$. По формуле Байеса получаем:

$$d_i(f) = \frac{1}{1-\beta_i(f)} \sum_{\nu=1}^{\mu} b_{i\nu}(f) g(\nu, f), \quad D_i(f) = \frac{1}{\beta_i(f)} \sum_{\nu=\mu+1}^{m} b_{i\nu}(f) g(\nu, f),$$

где $\beta_i(f) = b_{i,\mu+1}(f) + \ldots + b_{im}(f)$ — вероятность появления предпочтительного исхода или, то же самое, что вероятность поглощения цепи в одном из эргодических классах $E_{\mu+1}, \ \ldots, \ E_m$, исходящей из невозвратного состояния $i \in E_0$ при данной $f \in F$.

Чтобы определить предпочтительные решения в невозвратных состояниях $i \in E_0$, в формулах, определяющих функции полезностей $U_l(x), \ l = 1, 2, 3$, вместо величин r_1 и r_2 ставятся величины $d_i(f)$ и $D_i(f)$, соотвественно, и рассматривается отношение $\beta_i(f) \gtreqless U_l[g(0, f')]$. Если имеет место знак $>$ $(=, <)$, то $f \succ f'$ $(f \sim f', \ f \prec f')$.

Следует заметить, что при $r_1 = d_i(f), r_2 = D_i(f)$ и $x = g(0, f')$ отношение (3) принимает вид $g(0, f') \gtreqless g_i(f)$. Это значит, что функция полезности $U_1(x)$ при данных значениях по существу представляет собой критерий оптимальности $G(f^\infty)$.

8. Функции полезности $U_l(x), \ l = 1, 2, 3$ используем в примере 2. Результаты вычислений сведены в следующей таблице:

k	$d_5(k)$	$D_5(k)$	$U_1(56)$	$U_2(56)$	$U_3(56)$	$\beta_5(k)$
1	12	96	0,5238	0,5121	0,5006	0,5
2	14	96	0,5122	0,5061	0,5001	0,5
1	13,33	118,86	0,4043	0,4496	0,4908	0,4375
1	9,6	69,82	0,7705	0,6613	0,5794	0,6875

Эти результаты показывают, что все три функции полезности предпочтение отдают рискованному решению. Но для этой цели функция $U_1(x)$ рекомендует принять решения $3 \in F_5$. В то же время, обе другие функции $U_2(x)$ и $U_3(x)$ это решение отклоняют ($U_3(x)$ более уверенна, чем $U_2(x)$), как непригодное для риска и для этой цели рекомедуют принять решения $4 \in F_5$ ($U_2(x)$ более уверенна, чем $U_3(x)$).

9. Итак, традиционный критерий оптимальности управляемых марковских цепей — предельный средний доход, в случае нескольких эргодических классов и невозвратных состояний может привести к неприемлемим решениям. Предпочтительные решения могут быть найдены с помощью функций полезностей, построенных на основе теории статистических решений и полезности.

Результаты данной работы могут быть распространены для марковских игр с несколькими эргодическими классами, изученные Ибрагимовым (2003).

Автор надеется, что соединение теорий управляемых марковских процессов и статистических решений и полезности не только расширяет области их приложения, но и способствует их развитию.

Список литературы

[1] Майн Х., Осаки С. Марковские процессы принятия решений. М.: Наука, 1977.

[2] Ибрагимов А.А. Построение функции полезности для статистической задачи решения с входной платой // Узбекский журнал "Проблемы информатики и энергетики". 1997, №5. С. 3-8.

[3] Де Грот М. Оптимальные статистические решения М.: Мир, 1974.

[4] Фишберн П. Теория полезности для принятия решений. М.: Наука, 1978.

[5] Ибрагимов А.А. Марковские игры с несколькими эргодическими классами // Украинский математический журнал. 2003, Т.55, №6. С. 466-471.