ОЦЕНКА РИСКА ПРОГНОЗИРОВАНИЯ СТАЦИОНАРНЫХ ЦЕПЕЙ МАРКОВА

К.С. Матецкий, Ю.С. Харин

Белорусский государственный университет,
кафедра математического моделирования и анализа данных
пр. Независимости, д. 4, г. Минск, Республика Беларусь
телефон: + (37529) 2642270; e-mail: kmateckiy@gmail.com

Представлены результаты исследования риска прогнозирования стационарных цепей Маркова для (0-1)-функции потерь. В случае неизвестной матрицы вероятностей одношаговых переходов получены теоретические и экспериментальные оценки риска.

Прогнозирование, риск, стационарная цепь Маркова

1 ВВЕДЕНИЕ

Цепи Маркова нашли применение в таких областях науки как экономика, техника, медицина, защита информации [1-3]. Одной из часто встречаемых на практике задач является прогнозирование будущих значений цепи Маркова. В данном докладе представлены результаты исследования риска прогнозирования стационарных цепей Маркова.

Будем рассматривать стационарную цепь Маркова x_t, $t \in N$ с конечным пространством состояний $A = \{0,1,\dots,N-1\}$, $N \geq 2$. Пусть $P \in R^{N \times N}$ – матрица вероятностей одношаговых переходов, $\pi \in R^N$ – ее стационарное распределение. По наблюдаемой последовательности $X_1^T = (x_1, x_2, \dots, x_T) \in A^T$ длительности $T \geq 2$ прогнозируются будущие значения $\hat{x}_{T+\tau}$, где $\tau \geq 1$ – горизонт прогнозирования. Задача состоит в исследовании риска прогнозирования для (0-1)-функции потерь.

2 РИСК ПРОГНОЗИРОВАНИЯ ЦЕПИ МАРКОВА ПРИ ИЗВЕСТНЫХ ПАРАМЕТРАХ МОДЕЛИ

В случае известных матрицы вероятностей одношаговых переходов P и стационарного распределения π оптимальная по критерию минимума риска прогнозирования для (0-1)-функции потерь прогнозирующая статистика имеет вид [1]:

$$\hat{x}_{T+\tau} = \arg\max_{j \in A} (P^\tau)_{x_T, j}. \tag{3.1}$$

Теорема 1. При известных параметрах стационарной цепи Маркова, риск для прогнозирующей статистики (3.1) равен

$$r_0(\tau) = 1 - \sum_{i \in A} \pi_i \max_{j \in A} (P^\tau)_{i,j}.$$

3 РИСК ПРОГНОЗИРОВАНИЯ ЦЕПИ МАРКОВА ПРИ НЕИЗВЕСТНЫХ ПАРАМЕТРАХ МОДЕЛИ

В том случае, когда матрица вероятностей одношаговых переходов не известна, строится ее оценка максимального правдоподобия:

$$\hat{p}_{i,j} = \begin{cases} \sum_{t=1}^{T-1} \delta_{x_t,i} \delta_{x_{t+1},j} / \sum_{t=1}^{T-1} \delta_{x_t,i}, & \text{если } \sum_{t=1}^{T-1} \delta_{x_t,i} \neq 0, \\ 1/N, & \text{если } \sum_{t=1}^{T-1} \delta_{x_t,i} = 0. \end{cases}$$

При этом «подстановочная» прогнозирующая статистика имеет вид:

$$\tilde{x}_{T+\tau} = \arg\max_{j \in A} (\hat{P}^\tau)_{x_T, j}. \tag{4.1}$$

В дальнейшем будем рассматривать случай $N = 2$, $\tau = 1$.

Теорема 2. Риск «подстановочной» прогнозирующей статистики (4.1) равен

$$r(T,1) = 1 - p_{01}\pi_0 - p_{10}\pi_1 - (p_{00} - p_{01})(\pi_0 S_1 + \pi_1 S_2) - (p_{11} - p_{10})(\pi_0 S_3 + \pi_1 S_4),$$

где

$$S_1 = p_{00}^{T-1} + \sum_{i=1}^{[(T-2)/3]} \sum_{j=2i+1}^{T-1-i} C_j^i C_{T-2-j}^{i-1} p_{00}^{j-i} p_{01}^i p_{10}^i p_{11}^{T-1-j-i} + \frac{1}{2} \sum_{i=1}^{[(T-1)/3]} C_{2i}^i C_{T-2-2i}^{i-1} p_{00}^i p_{01}^i p_{10}^i p_{11}^{T-1-3i},$$

$$S_2 = \sum_{i=0}^{[(T-3)/3]} \sum_{j=2i+1}^{T-2-i} C_j^i C_{T-2-j}^i p_{00}^{j-i} p_{01}^i p_{10}^{i+1} p_{11}^{T-2-j-i} + \frac{1}{2} \sum_{i=0}^{[(T-2)/3]} C_{2i}^i C_{T-2-2i}^i p_{00}^i p_{01}^i p_{10}^{i+1} p_{11}^{T-2-3i},$$

$$S_4 = p_{11}^{T-1} + \sum_{i=1}^{[(T-2)/3]} \sum_{j=2i+1}^{T-1-i} C_j^i C_{T-2-j}^{i-1} p_{00}^{j-i} p_{01}^i p_{10}^i p_{11}^{T-1-j-i} + \frac{1}{2} \sum_{i=0}^{[(T-2)/3]} C_{2i}^i C_{T-2-2i}^i p_{00}^{T-2-3i} p_{01}^{i+1} p_{10}^i p_{11}^i,$$

$$S_4 = p_{11}^{T-1} + \sum_{i=1}^{[(T-2)/3]} \sum_{j=2i+1}^{T-1-i} C_j^i C_{T-2-j}^{i-1} p_{00}^{T-1-j-i} p_{01}^i p_{10}^i p_{11}^{j-i} +$$
$$\frac{1}{2} \sum_{i=1}^{[(T-1)/3]} C_{2i}^i C_{T-2-2i}^{i-1} p_{00}^{T-1-3i} p_{01}^i p_{10}^i p_{11}^i.$$

Рассмотрим частный случай задачи, когда матрица вероятностей одношаговых переходов является симметрической:

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{01} & p_{00} \end{pmatrix},$$

где $p_{00} + p_{01} = 1$, $0 < p_{00} < 1$. Вектор стационарных вероятностей в этом случае равен

$$\pi = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}.$$

Теорема 3. В случае, когда матрица вероятностей одношаговых переходов симметрическая, риск «подстановочной» прогнозирующей статистики (4.1) равен

$$r(T,1) = r_0(1) + R(T),$$

где

$$r_0(1) = \min\{p_{00}, p_{01}\},$$
$$R(T) = (p_{01} - p_{00})(S(T) - I\{p_{00} > p_{01}\}),$$
$$S(T) = p_{00}^{T-1} + \sum_{i=1}^{[(T-2)/3]} p_{00}^{T-1-2i} p_{01}^{2i} \sum_{j=2i+1}^{T-1-i} C_j^i C_{T-2-j}^{i-1} +$$
$$\frac{1}{2} \sum_{i=1}^{[(T-1)/3]} C_{2i}^i C_{T-2-2i}^{i-1} p_{00}^{T-1-2i} p_{01}^{2i} +$$
$$\sum_{i=0}^{[(T-3)/3]} p_{00}^{T-2-2i} p_{01}^{2i+1} \sum_{j=2i+1}^{T-2-i} C_j^i C_{T-2-j}^i +$$
$$\frac{1}{2} \sum_{i=0}^{[(T-2)/3]} C_{2i}^i C_{T-2-2i}^i p_{00}^{T-2-2i} p_{01}^{2i+1}.$$

Теорема 4. Если параметры модели не известны, то риск прогнозирования статистики (4.1) удовлетворяет асимптотическому разложению

$$r(T,1) = \sum_{i,j \in A} \pi_i p_{i,j} \Phi\left(\sqrt{T-1} \frac{0.5 - p_{i,j}}{\sqrt{p_{i,j}(1-p_{i,j})/\pi_i}} \right) + O(\rho^T), \qquad (4.2)$$

где $\rho \in (0;1)$.

4 РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

На рис.1. приведены графики оценки риска прогнозирования методом Монте-Карло и главного члена асимптотического разложения (4.2) для цепи Маркова с матрицей вероятностей одношаговых переходов $P = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}$, для горизонта прогнозирования $\tau = 1$, при T, изменяющемся от 2 до 50. Количество итераций $K = 1000000$.

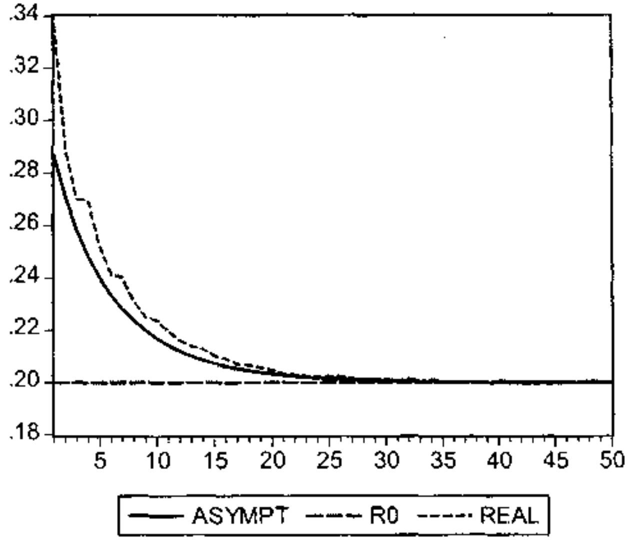

Рис.1. Риск прогнозирования.

ЛИТЕРАТУРА

[1] Харин, Ю.С. Оптимальность и робастность в статистическом прогнозировании / Ю.С. Харин — Мн.: БГУ, 2008.

[2] Basawa, I.V. Statistical interference for stochastic processes / I.V. Basawa, B.L.C. Rao. — N.Y.: Academic Press, 1980.

[3] Дуб, Дж. Вероятностные процессы / Дж. Дуб. — М.: ИЛ, 1956.