СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ О ПАРАМЕТРАХ ВЕКТОРНОЙ ЦЕПИ МАРКОВА С ЧАСТИЧНЫМИ СВЯЗЯМИ

Н. С. Сологуб

ВВЕДЕНИЕ

При математическом моделировании сложных систем и процессов в экономике, технике, медицине, социологии, генетике и других приложениях часто возникает необходимость построения адекватных вероятностно-статистических моделей дискретных временных рядов $x_t \in V$, $t \in N$, где пространство состояний $V = \{0,1,\dots,N-1\}$ – конечное множество мощности $N \geq 2$ с длинной памятью [1]. Известной моделью таких дискретных временных рядов является цепь Маркова достаточно высокого порядка $s \in N$, определяющего длину памяти; если $s = 1$, то цепь Маркова называется простой, если $s > 1$ – сложной [2]. Однако для такой модели число параметров D растет экспоненциально при увеличении порядка s: $D = N^s(N-1)$, и для статистического оценивания параметров требуется иметь реализацию $x_1,\dots,x_T$ не всегда доступной на практике длительности $T > D$. В связи с этим актуальна проблема построения малопараметрических моделей цепей Маркова высокого порядка. Примерами таких моделей являются цепь Маркова с частичными связями, модель Рафтери, цепь Маркова условного порядка, цепь Маркова переменного порядка [3, 4].

Отметим также, что векторная цепь Маркова с частичными связями не может рассматриваться как семейство одномерных цепей Маркова из-за наличия зависимости между компонентами вектора в модели $VMC(s, M_r)$.

ВЕКТОРНАЯ ЦЕПЬ МАРКОВА С ЧАСТИЧНЫМИ СВЯЗЯМИ

Назовем модель векторной цепью Маркова с частичными связями $VMC(s, M_r)$, если распределение вероятностей наблюдаемого вектора в момент времени $t+1$ зависит не от всех ms компонент s предшествующих векторов, а только от конкретных r компонент, задаваемых шаблон-множеством M_r. Пусть $J = (J_1,\dots J_s)$ – матрица s-предыстории размера $m \times s$; $M_r = (M_{ij})$ - бинарная матрица размера $m \times s$, причем $M_{ij} = 1$, если условное распределение вероятностей (8) вектора x_{s+1} зависит от x_{ij} (iой компоненты j-ого вектора), и $M_{ij}=0$ иначе. При этом $r=\sum_{i=1}^{m}\sum_{j=1}^{s}M_{ij}$ – число связей, $r\in\{1,2,\ldots,m\cdot s\}$. Если $r=ms$, то описанная модель представляет собой полносвязную векторную цепь Маркова порядка s.

Теперь определим матрицу вероятностей одношаговых переходов. Пусть $\alpha_1,\ldots,\alpha_r$ – значения тех r компонент матрицы J, от которых зависит распределение вероятностей будущего вектора x_{t+1}. Тогда

$$p_{J;J_{s+1}}(t)=q_{(\alpha_1,\ldots,\alpha_r),J_{s+1}}(t),\qquad \alpha_i\in V. \tag{1}$$

Для матрицы $Q=(q_{(\alpha_1,\ldots,\alpha_r),J}(t))$ должно выполняться условие нормировки:

$$\sum_{J\in V^m}q_{(\alpha_1,\ldots,\alpha_r),J}(t)=1,\quad \alpha_1,\ldots,\alpha_r\in V,J\in V^m. \tag{2}$$

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ О ПАРАМЕТРАХ МОДЕЛИ

Следуя [4], построим статистический критерий для проверки гипотезы о параметрах $Q=(q_{\alpha,J})$ модели $HVMC(s,M_r)$ на основе статистики

$$\rho=\sum_{\alpha\in V^r}\sum_{J\in V^m}\frac{m(\alpha)}{q_{\alpha,J}}\left(\hat{q}_{\alpha,J}-q_{\alpha,J}\right)^2, \tag{3}$$

где:

$m(\alpha)$ – количество таких матриц размера m × s в реализации, что состояние шаблона равно α;

$q_{\alpha,J}$ – элемент истинной матрицы переходов Q;

$\hat{q}_{\alpha,J}$ – элемент оценки максимального правдоподобия $\hat{Q}$ матрицы Q.

Теорема 1. Пусть для реализации $HVMC(s,M_r)$ известен истинный шаблон M_r. Тогда при $n\to\infty$ распределение статистики ρ из (1) сходится по вероятости к χ^2 распределению с $L=N^r(N^m-1)$ степенями свободы.

Теорема 2. Если

$$H_{1T}:q_{1,(\alpha,J)}=q_{0,(\alpha,J)}+\frac{c_{(\alpha,J)}}{\sqrt{T-s}}+O(1/T),\quad T\to\infty\ – \tag{4}$$

семейство контигуальных альтернатив, где константы $c_{(\alpha,J)}:\alpha\in V^r,J\in V^m$ таковы, что

$$\sum_{J\in V^m}c_{(\alpha,J)}=0,\qquad \sum_{\alpha\in V^r,J\in V^m}c^2_{(\alpha,J)}>0,$$

то при $\Delta = G_L^{-1}(1-\varepsilon)$ асимптотическая мощность теста (3):

$$w \to 1 - G_{L,a}(G_L^{-1}(1-\varepsilon)), \quad (5)$$

где $G_{L,a}(\cdot)$ – функция нецентрального распределения χ^2 с L степенями свободы и параметром нецентральности

$$a = \sum_{\alpha \in V^r, J \in V^m} \frac{m(\alpha) c^2_{(\alpha,J)}}{q_{0,(\alpha,J)}}. \quad (6)$$

Пусть шаблон связей M_r, размер шаблона r и порядок цепи Маркова s неизвестны. Пусть $M^a_{r_a}$ – действительный шаблон реализации; $M^e_{r_e}$ – оценка шаблона; α_a – состояние истинного шаблона $M^a_{r_a}$; α_a – состояние тех компонент оценки шаблона $M^e_{r_e}$, которые не пересекаются с компонентами истинного шаблона $M^a_{r_a}$; $m(\alpha_a, \alpha_e)$ – количество таких отрезков реализации длины $\max\{s_a, s_e\}$, что состояние реального шаблона $M^a_{r_a}$ равно α_a, а состояние оценки шаблона $M^e_{r_e}$ равно α_a. Вычислим статистику ρ с учетом этих обозначений:

$$\rho = \sum_{\alpha_a \in V^{r_a}} \sum_{\alpha_e \in V^{r_e}} \sum_{J \in V^m} \frac{m(\alpha_a, \alpha_e)}{q_{\alpha_a,J}} \left(\hat{q}_{\alpha_e,J} - q_{\alpha_a,J} \right)^2. \quad (7)$$

Теорема 3. Построим критерий проверки гипотез $H_0 : Q = Q_0, M_r = M^e_{r_e}, r = r_e, s = s_e$ против альтернативы общего вида $H_1 : \overline{H_0}$, где Q_0 – некоторая фиксированная стохастическая матрица, задающая условные вероятности переходов для шаблона $M^e_{r_e}$. Решающее правило для проверки H_0 против альтернативы H_1 состоит в том, что принимается гипотеза

$$\begin{cases} H_0, & \rho \leqslant \Delta; \\ H_1, & \rho > \Delta. \end{cases} \quad (8)$$

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ О ПАРАМЕТРАХ МОДЕЛИ С НЕЗАВИСИМЫМИ КОМПОНЕНТАМИ

Модель векторной цепи Маркова с частичными связями и независимыми компонентами определяют m наборов параметров: порядок цепи s_i, количество связей r_i, шаблон связей $M_{r_i}^{(i)}$ и матрица одношаговых переходов $Q^{(i)}$. Определим гипотезы:

$$\begin{cases} H_0^{(i)} : q_{\alpha,J}^{(i)} = q_{0,(\alpha,J)}^{(i)}, \\ H_{1T}^{(i)} : q_{\alpha,J}^{(i)} = q_{0,(\alpha,J)}^{(i)} + \dfrac{c_{(\alpha,J)}^{(i)}}{\sqrt{T-s}} + O(1/T), \quad T \to \infty, \end{cases} \tag{9}$$

где $\{Q_0^{(i)}\}_{i=1}^{m}$ – набор m фиксированных стохастических матриц.

Теорема 4. Решающие правила для проверки гипотез о матрице одношаговых переходов для каждой компоненты модели $IVMC(s,\{M_{r_i}\}_{i=1}^{m})$ состоят в том, что принимается гипотеза

$$\begin{cases} H_0^{(i)} : q_{\alpha,J}^{(i)} = q_{0,(\alpha,J)}^{(i)}, \quad \rho_i \leqslant \Delta; \\ H_{1T}^{(i)} : q_{\alpha,J}^{(i)} = q_{0,(\alpha,J)}^{(i)} + \dfrac{c_{(\alpha,J)}^{(i)}}{\sqrt{T-s}} + O(1/T), \quad \rho_i > \Delta. \end{cases} \tag{10}$$

Асимптотическая оценка мощности построенных тестов имеет вид:

$$w_i \to 1 - G_{L_i,a_i}(G_{L_i}^{-1}(1-\varepsilon)), \tag{11}$$

где $G_{L_i,a_i}(\cdot)$ – функция нецентрального распределения χ^2 с $L_i = N^{r_i}(N-1)$ степенями свободы и параметром нецентральности

$$a_i = \sum_{\alpha \in V^{r_i}, J \in V} \frac{m(\alpha)\left(c_{(\alpha,J)}^{(i)}\right)^2}{q_{0,(\alpha,J)}^{(i)}}. \tag{12}$$

Литература

1. *Харин, Ю. С.* Векторная цепь Маркова с частичными связями и статистические выводы о ее параметрах / Ю. С. Харин, М. В. Мальцев, Н.С. Сологуб // Доклады НАН Беларуси. – 2016. – № 1. – С. 34–43
2. *Харин, Ю.С.* Цепи Маркова с r-частичными связями и их статистическое оценивание // Доклады НАН Беларуси. – 2004. – Т. 48, № 1. –С. 40–44.
3. *Харин, Ю.С.,* Вероятностно-статистический анализ цепей Маркова высокого порядка // Вестник БГУ. Сер. 1. 2006. № 3
4. *Kemeny, J. G.* Finite Markov chains / J. G. Kemeny, J. L. Snell . – Princeton : Van Nostrand, 1960. – 525 c.