

АВТОРЕГРЕССИОННЫЕ ВРЕМЕННЫЕ РЯДЫ ПРИ НАЛИЧИИ КЛАССИФИКАЦИИ НАБЛЮДЕНИЙ

А. В. Рудаковская

ВВЕДЕНИЕ

Модель авторегрессионных временных рядов часто встречается в статистике. При помощи модели авторегрессионного временного ряда описываются процессы и явления в различных сферах: в экономике и в промышленности [1], в медицине [2] и др.

Одним из типов искажений авторегрессионных временных рядов является классификация наблюдений – регистрация вместо истинного значения временного ряда лишь номера класса, в который попало исходное наблюдение. Искажения такого типа описаны в [3, 4].

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть на (Ω, F, P) определен авторегрессионный временной ряд

$$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \dots + \theta_p x_{t-p} + \xi_t, t \in \mathbb{Z}, \quad (1)$$

где $p \in \mathbb{N}$ – порядок авторегрессии, $\theta = (\theta_i) \in \mathbb{R}^p$ – вектор коэффициентов авторегрессии, $\theta_p \neq 0$, $\{\xi_t\}$ – дискретный «белый шум», $\mathcal{L}\{\xi_t\} = N_1(0, \sigma^2)$. Временной ряд вида (1) называют временным рядом авторегрессии порядка p (краткое обозначение $AR(p)$).

Пусть задано борелевское разбиение числовой прямой на $2 \leq L < +\infty$ числовых промежутков:

$$\mathbb{R} = \bigcup_{i=0}^{L-1} A_{i+1}, \quad (2)$$

$$A_{i+1} = (a_i; a_{i+1}], -\infty = a_0 < a_1 < \dots < a_{L-1} < a_L = +\infty.$$

Вместо исходного временного ряда x_t наблюдаем классифицированный временной ряд y_t ,

$$y_t = \sum_{i=1}^L i I_{A_i}(x_t), \quad (3)$$

где y_t – номер класса, в который попало значение x_t ($I_A(x)$ – индикаторная функция множества A).

Полагаем известными порядок авторегрессии p , значения классифицированного временного ряда y_t , $t = \overline{0, T}$, а также интервалы $(a_i; a_{i+1}], i = \overline{0, L-1}$, по которым проводилась классификация.

Ставится задача оценивания неизвестных параметров $\theta_1, \theta_2, \dots, \theta_p$.

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ

Теорема 1. В данных условиях эксперимента классифицированный временной ряд y_t имеет дискретное распределение. Количество значений равно количеству интервалов классификации, при этом $p_i = P\left(y_t = i - \left[\frac{L+1}{2}\right]\right) = P(x_t \in [a_{i-1}; a_i)) = \Phi\left(\frac{a_i}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^p \theta_i^2}}}\right) - \Phi\left(\frac{a_{i-1}}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^p \theta_i^2}}}\right)$,

где Φ – функция распределения стандартного нормального закона.

Исследуем оценку, полученную по методу наименьших квадратов (4):

$$\begin{cases} (y_{-1}^2 + \dots + y_{T-1}^2)\theta_1 + (y_{-2}y_{-1} + \dots + y_{T-2}y_{T-1})\theta_2 + \dots + \\ + (y_{-p}y_{-1} + \dots + y_{T-p}y_{T-1})\theta_p = y_0y_{-1} + \dots + y_Ty_{T-1}; \\ (y_{-1}y_{-2} + \dots + y_{T-1}y_{T-2})\theta_1 + (y_{-2}^2 + \dots + y_{T-2}^2)\theta_2 + \dots + \\ + (y_{-p}y_{-2} + \dots + y_{T-p}y_{T-2})\theta_p = y_0y_{-2} + \dots + y_Ty_{T-2}; \\ \vdots \\ (y_{-1}y_{-p} + \dots + y_{T-1}y_{T-p})\theta_1 + (y_{-2}y_{-p} + \dots + y_{T-2}y_{T-p})\theta_2 + \dots + \\ + (y_{-p}^2 + \dots + y_{T-p}^2)\theta_p = y_0y_{-p} + \dots + y_Ty_{T-p}. \end{cases} \quad (4)$$

Замечание.

В дальнейшем будем исследовать авторегрессию первого порядка ($p = 1$). В этом случае получаем следующий процесс:

$$x_t = \theta x_{t-1} + \xi_t; \quad \mathcal{L}\{\xi_t\} = N_1(0, \sigma^2); \quad -1 < \theta < 1; \quad (5)$$

$$y_t = \text{sign}(x_t) = \begin{cases} 1, & x > 0; \\ -1, & x \leq 0. \end{cases} \quad (6)$$

СМЕЩЕНИЕ И ВАРИАЦИЯ ОЦЕНКИ. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

На рис. 1 приведены примеры графиков для значения смещения и вариации оценки в зависимости от длины реализации T при $\theta = 0.5$:

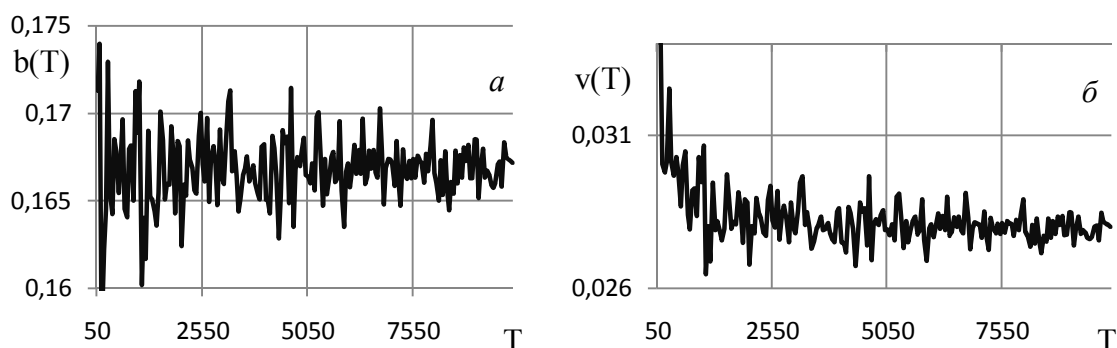


Рис. 1. Смещение оценки и вариация оценки

Здесь численные эксперименты проводились для различных длин реализаций временного ряда (от 50 до 10000), для каждого значения длины реализации проводилось 100 повторений экспериментов.

По этим результатам можно предположить, что оценка $\hat{\theta}$ является смещенной и величина смещения зависит от θ . Кроме того, с увеличением T видна сходимость вариации к некоторому константному значению.

СМЕЩЕНИЕ И ВАРИАЦИЯ ОЦЕНКИ. ТЕОРЕТИЧЕСКОЕ ОБОСНОВАНИЕ

Были доказаны следующие утверждения:

Лемма 1. Значения $p_{++} = P\{x_t > 0, x_{t+1} > 0\}$, $p_{+-} = P\{x_t > 0, x_{t+1} \leq 0\}$, $p_{-+} = P\{x_t \leq 0, x_{t+1} > 0\}$, $p_{--} = P\{x_t \leq 0, x_{t+1} \leq 0\}$ могут быть найдены при помощи формул (7) – (10).

$$p_{++}(\theta) = \sqrt{1 - \theta^2} \int_0^{+\infty} \varphi(\sqrt{1 - \theta^2}x) \Phi(\theta x) dx, \quad (7)$$

$$p_{+-}(\theta) = \sqrt{1 - \theta^2} \int_0^{+\infty} \varphi(\sqrt{1 - \theta^2}x) \Phi(-\theta x) dx; \quad (8)$$

$$p_{-+}(\theta) = \sqrt{1 - \theta^2} \int_0^{+\infty} \varphi(\sqrt{1 - \theta^2}x) \Phi(-\theta x) dx; \quad (9)$$

$$p_{--}(\theta) = \sqrt{1 - \theta^2} \int_{-\infty}^0 \varphi(\sqrt{1 - \theta^2}x) \Phi(-\theta x) dx; \quad (10)$$

где $\varphi(x)$, $\Phi(x)$ – плотность вероятности и функция распределения стандартного нормального закона соответственно.

Теорема 2. Оценка $\hat{\theta}$, построенная по методу наименьших квадратов для модели (5), (6) является асимптотически смещенной, причем смещение является функцией параметра θ и верно предельное соотношение

$$\lim_{T \rightarrow \infty} b(\hat{\theta}) = p_{++}(\theta) + p_{+-}(\theta) - p_{-+}(\theta) - p_{--}(\theta) - \theta, \quad (11)$$

где $p_{++}, p_{+-}, p_{-+}, p_{--}$ – выражения, задаваемые (7) – (10).

Для нахождения значений вариации оценки имеет место теорема

Теорема 3. МНК-оценка параметра θ модели (5), (6) имеет вариацию, асимптотически близкую к квадрату смещения оценки:

$$\lim_{T \rightarrow \infty} v(\hat{\theta}) = (b(\hat{\theta}))^2. \quad (12)$$

На рис. 2 представлены значения смещения и вариации, полученные численно и аналитически, в зависимости от значения параметра θ :

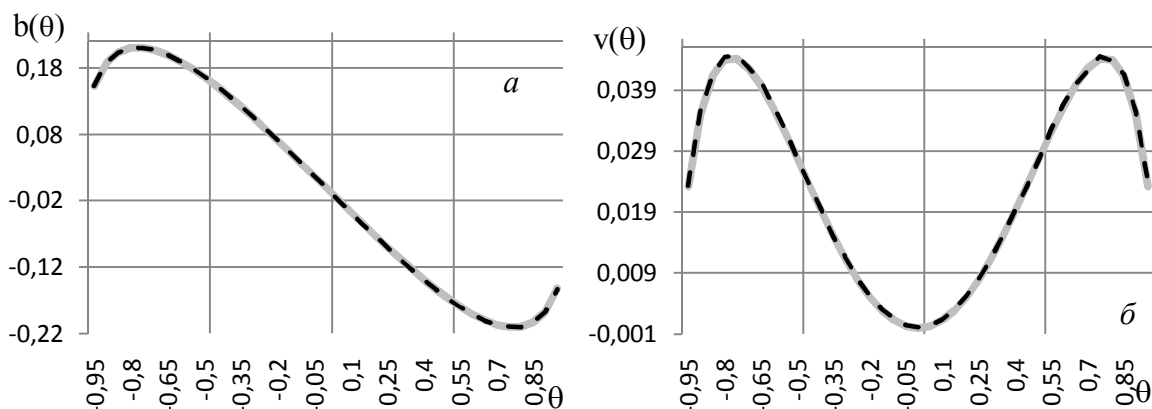


Рис. 2. Смещение и вариация оценки в зависимости от значения θ .

График серого цвета – теоретически полученная кривая,
черный пунктир – практически полученные значения

ЗАКЛЮЧЕНИЕ

В работе рассмотрена модель авторегрессионных временных рядов при наличии искажений – классификации наблюдений. Установлена связь между распределениями исходного и классифицированного рядов. Для неизвестных параметров θ построены оценки по методу наименьших квадратов. В случае авторегрессии первого порядка ($p = 1$) с классификацией значений на положительные-отрицательные изучены свойства этой оценки. В асимптотике оценка оказалась смещенной относительно истинного значения параметра θ на некоторую величину, зависящую от θ , причем зависимость не является линейной. Вариация оценки близка к квадрату смещения.

Литература

1. Бокс ДЖ., Дженкинс Г. Анализ временных рядов, прогноз и управление / М.: Мир. 1974.
2. Weber R. 8 lectures to Cambridge M.Phil in Statistics students, Complete Notes / Cambridge, 1999.
3. Агеева Е. С. Статистическое оценивание параметров множественной регрессии при наличии классификации наблюдений // Сборник материалов республиканской научной конференции НИРС-2011, Минск, 2011, с. 72.
4. Харин Ю. С. Оптимальность и робастность в статистическом прогнозировании. Мн., БГУ.