АНАЛИЗ ПЕРИОДИЧЕСКИХ АВТОРЕГРЕССИЙ

Ю. В. Меленец

Белорусский государственный университет
г. Минск, Беларусь
E-mail: melene@tut.by

. Рассматривается задача оценивания параметров и восстановления пропущенных значений нестационарных случайных процессов с периодическими вероятностными характеристиками в ситуации, когда процессы описываются авторегрессионными уравнениями с периодическими коэффициентами. Приводятся результаты имитационного моделирования

Ключевые слова: период, авторегрессия, оценки.

Пусть $\{X_t$, t=1, 2, …} – нестационарный случайный процесс, математическое ожидание и ковариации которого ограничены и являются периодическими функциями. Возможность описания такого процесса уравнениями авторегрессии с постоянными коэффициентами весьма проблематична [1]. В связи с этим предлагается рассматривать модель авторегрессии, параметры которой являются периодическими функциями:

$$X_t = \sum_{k=1}^{n} a_k(t) X_{t-k} + \xi_t \quad , \tag{1}$$

где $\xi_t, t = 1, 2, ...$ – нестационарная последовательность независимых случайных величин с $M\xi_t = 0$, $D\xi_t = \sigma_t^2$, $\sigma_\tau^2 = \sigma_{\tau+jT}^2$; $a_k(\tau) = a_k(\tau + jT)$, $1 \le k \le n$, $1 \le \tau \le T$, $j = 1, 2, ...$; T – натуральное число, имеющее смысл периода.

Вопросом исследования свойств временного ряда X_t, получаемого при помощи (1), посвящен ряд работ [2-4]. В настоящей статье предполагается, что период T известен, и рассматриваются две задачи.

1. В ситуации, когда ξ_t есть гауссовская последовательность, необходимо по наблюдениям $x_1, x_2, ..., x_N$ оценить nT коэффициентов $a_k(\tau)$, $1 \le k \le n$, $1 \le \tau \le T$, T дисперсий σ_τ^2, $1 \le \tau \le T$ и исследовать свойства получаемых оценок.
2. Построить оценку одного из пропущенных значений в выборке $x_1, x_2, ..., x_N$ по известным значениям X_t до и после пропуска и исследовать свойства этой оценки.

Отметим, что задача оценивания коэффициентов $a_k(\tau)$ в модели (1) рассматривалась в [2]. Здесь использовалась система уравнений, аналогичных уравнениям Юла-Уолкера для стационарных авторегрессий с постоянными коэффициентами. Замена теоретических ковариаций на выборочные и решение получившейся системы относительно коэффициентов $a_k(\tau)$ давало оценки этих параметров. Но практическая реализация этой процедуры малоэффективна и требует асимптотически больших объемов выборки.

Для решения первой задачи применим метод максимального правдоподобия. Первые n элементов выборки $x_1, x_2, ..., x_n$ считаем начальными значениями авторегрессии. Запишем плотность распределения случайных величин $\xi_{n+1}, \xi_{n+2}, ..., \xi_N$ в виде:

$$(2\pi)^{\frac{n-N}{2}} \left(\prod_{t=n+1}^{N} \sigma_t \right)^{-1} \exp\left\{ -\frac{1}{2} \sum_{t=n+1}^{N} \sigma_t^{-2} \xi_t^2 \right\}, \tag{2}$$

и перейдем в (2) от переменных $\xi_{n+1}, \xi_{n+2}, ..., \xi_N$ к переменным $x_{n+1}, x_{n+2}, ..., x_N$. Из (1) получаем, что якобиан этого преобразования равен 1, и совместная плотность значений $x_{n+1}, x_{n+2}, ..., x_N$ есть:

$$f\left(X_{n+1} = x_{n+1}, ..., X_N = x_n / \left\{ a_k(\tau), \sigma_\tau^2, 1 \le \tau \le T, 1 \le k \le n \right\} \right) =$$
$$= (2\pi)^{\frac{n-N}{2}} \left(\prod_{t=n+1}^{N} \sigma_t \right)^{-1} \exp\left\{ -\frac{1}{2} \sum_{t=n+1}^{N} \sigma_t^{-2} \left(x_t - \sum_{k=1}^{n} a_k(t) x_{t-k} \right)^2 \right\}. \tag{3}$$

Введем обозначения: $\alpha_T = \left[\dfrac{N-n}{T} \right]$ - целое число циклов, содержащихся в наблюдаемой реализации $x_{n+1}, x_{n+2}, ..., x_N$; $N - n = \alpha_T T + \beta_T$, $0 \le \beta_T < T$;

$$m_t = \begin{cases} \alpha_T + 1, если 1 \le t \le \beta_T \\ \alpha_T, если \beta_T < t \le T \end{cases}$$

Из (3) и введенных обозначений следует, что логарифм функции правдоподобия оцениваемых параметров записывается в виде:

$$l\left(\left\{ a_k(\tau), \sigma_\tau^2, 1 \le \tau \le T, 1 \le k \le n \right\} \right) = \frac{n-N}{2} \ln 2\pi - \sum_{\tau=1}^{T} \sum_{j=1}^{m_\tau} \sigma_{n+\tau+(j-1)T} -$$
$$-\frac{1}{2} \sum_{\tau=1}^{T} \sum_{j=1}^{m_\tau} \sigma_{n+\tau+(j-1)T}^{-2} \left(x_{n+\tau+(j-1)T} - \sum_{k=1}^{n} a_k\left(n+\tau+(j-1)T \right) * x_{n+\tau+(j-1)T-k} \right)^2. \tag{4}$$

Используя (4) и свойства периодичности параметров $\left\{ a_k(\tau), \sigma_\tau^2 \right\}$, получаем, что задача оценивания совокупности этих параметров распадается на T однотипных задач. При фиксированном значении τ, $1 \le \tau \le T$ оценки коэффициентов $a_k(\tau)$, $1 \le k \le n$ находятся как решение задачи:

$$\sum_{j=1}^{m_\tau} \left(x_{n+\tau+(j-1)T} - \sum_{k=1}^{n} a_k\left(n+\tau \right) * x_{n+\tau+(j-1)T-k} \right)^2 \to \min_{\{a_k(\tau)\}}, \tag{5}$$

а дисперсия $\sigma_{\tau+n}^2$ по найденным оценкам $\hat{a}_k(\tau)$ определяется по следующей формуле:

$$\widehat{\sigma^2_{\tau+n}} = \frac{1}{m_\tau - 1} \sum_{j=1}^{m_\tau} \left(x_{n+\tau+(j-1)T} - \sum_{k=1}^{n} \widehat{a_k}\left(n+\tau \right) * x_{n+\tau+(j-1)T-k} \right)^2. \tag{6}$$

Решение задачи (5) записывается в виде:

$\hat{A}_\tau = \left(Z_\tau^* Z_\tau \right)^{-1} Z_\tau^* X_\tau$,

где обозначено: $\hat{A}_\tau = \left[\hat{a}_k(n+\tau), 1 \le k \le n \right]^*$, $X_\tau = \left[x_{n+\tau+(j-1)T}, 1 \le j \le m_\tau \right]^*$, $Z_\tau = \left\| x_{n+(i-1)T+\tau-j}, 1 \le i \le m_\tau, 1 \le j \le n \right\|$.

Для исследования свойств оценок (6) и (7) введем в рассмотрение матрицу $B(r) = \left\| b_{ij}(r) \right\|$, $1 \le i, j \le T$, где

$$b_{ij} = \begin{cases} r^n, i = j, \\ -a_{i-j}(i) r^{n-(i-j)}, i > j, \\ -a_{T+i-j}(i) r^{n-(T+i-j)}, i < j, \end{cases}$$

и предположим, что корни уравнения det*B(r)*=0 ограничены по модулю единицей. Для этого случая в [3] показано, что

$$X_t = \sum_{j=0}^{\infty} \delta_j(t) \xi_{t-j}, \tag{8}$$

где коэффициенты $\delta_j(t)$ периодичны по t с периодом T и ряд в правой части (8) сходится в среднеквадратическом смысле. Пи выполнении (8) удается доказать, что оценки (6) и (7) являются состоятельными. Доказательство проводится аналогично тому, как это делается в [5] для стационарных авторегрессий с постоянными коэффициентами, и не проводится из-за громоздкости.

Алгоритм оценивания параметров периодических авторегрессий по методу максимального правдоподобия реализован на ЭВМ, и исследование возможности его применения проводилось методами статистического моделирования. Приведем в качестве примера результаты просчетов для авторегрессии порядка *n*=1 с периодом *T*=5 и с объемами выборок *N*=500 и *N*=1000. обработке подвергалось по 50 реализаций. Значения параметров, их оценки и доверительные интервалы с уровнем значимости $\alpha = 0,05$ приведены в таблице, анализ данных которой свидетельствует о работоспособности рассматриваемого алгоритма.

Во второй задаче предполагается, что в выборке $x_1, x_2, ..., x_N$ имеется $\left[l, l+s \right]$ – интервал пропущенных значений, $1 \le l \le N - s$, и требует восстановить одно из этих значений, относящихся к моменту $l+m$, $0 \le m \le s$. Интерполяционную формулу для оценки представим в виде:

$$\hat{x}_{l+m} = \sum_{j=1}^{N_1} d_j x_{l-j} + \sum_{j=1}^{N_3} h_j x_{l+s+j}, \tag{9}$$

где $l \le N_1 \le l-1$, $1 \le N_2 \le N - s - l$. Качество ошибки (9) будем характеризовать средним квадратом оценки: $M\left(x_{l+m} - \hat{x}_{l+m} \right)^2$, то есть параметры $\left\{ d_1, ..., d_{N_1}, h_1, ..., h_{N_2} \right\}$ должны определяться как решения задачи:

$$M\left(x_{l+m} - \sum_{j=1}^{N_1} d_j x_{l-j} - \sum_{j=1}^{N_3} h_j x_{l+s+j} \right)^2 \to \min_{\{d_j, h_j\}}. \tag{10}$$

Обозначим $R_\tau(k) = M(x_\tau x_{\tau+k})$ – ковариации длины k процесса x_t в точке $t = \tau$. Из (10) следует, что искомые коэффициенты $\left\{ d_j, h_j \right\}$ и удовлетворяют следующей системе линейных алгебраических уравнений:

$$\begin{gathered} \sum_{j=1}^{N_1} d_j R_l(j-k) + \sum_{j=1}^{N_2} h_j R_l(s+j+k) = R_l(m+k), 1 \le k \le N_1 \\ \sum_{j=1}^{N_1} d_j R_l(s+j+k) + \sum_{j=1}^{N_2} h_j R_l(j-k) = R_l(m-s-k), 1 \le k \le N_2, \end{gathered} \tag{11}$$

решение которой не предоставляет труда. Если коэффициенты $\{a_k(\tau), 1 \le k \le n, 1 \le \tau \le T\}$ и дисперсии $\{\sigma_\tau^2, 1 \le \tau \le T\}$ известны, то ковариации $R_l(k)$ процесса X_t из (1) могут быть найдены из системы уравнений Юла-Уолкера [2], в противном случае теоретические ковариации могут быть заменены их выборочными оценками.

Численные эксперименты восстановления пропущенных значений для модели (1) реализованы на ЭВМ и их работоспособность доказана методами имитационного моделирования.