

# 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.

При обработке временных рядов часто выясняется, что в наблюдаемых реализациях проявляются более или менее регулярные колебания. В этом случае в качестве математической модели временного ряда выбирают, как правило, модель периодического тренда и случайной помехи. Задача оценивания параметров сводится к задаче оценивания коэффициентов в разложении тренда на периодические компоненты. Но в такой модели наблюдения являются независимыми, а это есть сильное ограничение, которое необходимо преодолевать. Поэтому предлагается другой подход в описании таких временных рядов.

Пусть  $\{X_t, t=1, 2, \dots\}$  – 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:

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

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

Вопросом исследования свойств временного ряда  $X_t$ , получаемого при помощи (1), посвящен ряд работ [2-3]. В настоящей статье предполагается, что период  $T$  известен, и рассматриваются две задачи.

1. В ситуации, когда  $\xi_t$  есть гауссовская последовательность, необходимо по наблюдениям  $x_1, x_2, \dots, x_N$  оценить  $nT$  коэффициентов  $a_k(\tau), 1 \leq k \leq n, 1 \leq \tau \leq T, T$  дисперсий  $\sigma_\tau^2, 1 \leq \tau \leq T$  и исследовать свойства получаемых оценок.

2. Построить оценку одного из пропущенных значений в выборке  $x_1, x_2, \dots, x_N$  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  $x_1, x_2, \dots, x_n$  считаем начальными значениями авторегрессии. Запишем плотность распределения случайных величин  $\xi_{n+1}, \xi_{n+2}, \dots, \xi_N$  в виде:

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

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

$$\begin{aligned} f \left( X_{n+1} = x_{n+1}, \dots, X_N = x_N / \{ a_k(\tau), \sigma_\tau^2, 1 \leq \tau \leq T, 1 \leq k \leq n \} \right) = \\ = (2\pi)^{\frac{n-N}{2}} \left( \prod_{t=n+1}^N \sigma_t \right)^{-1} \exp \left\{ -\frac{1}{2} \sum_{t=n+1}^N \sigma_t^{-2} \left( x_t - \sum_{k=1}^n a_k(t) x_{t-k} \right)^2 \right\}. \end{aligned} \quad (3)$$

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

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

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

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

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

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

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

$$\hat{\sigma}_{\tau+n}^2 = \frac{1}{m_\tau - 1} \sum_{j=1}^{m_\tau} \left( x_{n+\tau+(j-1)T} - \sum_{k=1}^n \hat{a}_k(n+\tau) x_{n+\tau+(j-1)T-k} \right)^2. \quad (6)$$

Решение задачи (5) записывается в виде:

$$A_\tau = (Z_\tau^* Z_\tau)^{-1} Z_\tau^* X_\tau,$$

где обозначено:

$$A_\tau = \left[ \hat{a}_k(n+\tau), 1 \leq k \leq n \right]^*, \quad X_\tau = \left[ x_{n+\tau+(j-1)T}, 1 \leq j \leq m_\tau \right]^*, \quad Z_\tau = \left\| x_{n+(i-1)T+\tau-j}, 1 \leq i \leq m_\tau, 1 \leq j \leq n \right\|.$$

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

$$X_i = \sum_{j=0}^{\infty} \delta_j(t) \xi_{i-j}, \quad (8)$$

где коэффициенты  $\delta_j(t)$  периодичны по  $t$  с периодом  $T$  и ряд в правой части (8) сходится в среднеквадратическом смысле. При выполнении (8) удастся доказать, что оценки (6) и (7) являются состоятельными. Доказательство проводится аналогично тому, как это делается в [4] для стационарных авторегрессий с постоянными коэффициентами..

Алгоритм оценивания параметров периодических авторегрессий по методу максимального правдоподобия реализован на ПК, 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 1, анализ данных которой свидетельствует о работоспособности рассматриваемого алгоритма.

Таблица 1

Параметры	N=500		N=1000	
	Оценки	Доверительные интервалы	Оценки	Доверительные интервалы
$a_1(1) = 0,2$	0,2026	[0,1976, 0,2076]	0,1989	[0,1958, 0,2019]
$a_1(2) = 0,8$	0,7906	[0,7856, 0,7956]	0,7942	[0,7913, 0,7970]
$a_1(3) = -1,5$	-1,4935	[-1,5141, -1,4729]	-1,5075	[-1,5577, -1,4974]
$a_1(4) = 1,2$	1,1936	[1,1727, 1,2145]	1,2015	[1,1853, 1,2177]
$a_1(5) = 2,0$	1,9888	[1,9731, 2,0045]	2,0087	[1,9986, 2,0188]
$\sigma_1^2 = 0,5$	0,4698	[0,4584, 0,4812]	0,4948	[0,4890, 0,5006]
$\sigma_2^2 = 1,0$	0,9584	[0,9331, 0,9827]	1,0213	[0,9203, 1,1223]
$\sigma_3^2 = 0,8$	0,8314	[0,8114, 0,8514]	0,7926	[0,7830, 0,8022]
$\sigma_4^2 = 3,5$	3,7001	[3,6670, 3,7232]	3,5206	[3,5100, 3,5312]
$\sigma_5^2 = 2,4$	2,2602	[2,2498, 2,2706]	2,4013	[2,3957, 2,4069]

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

$$\hat{x}_{l+m} = \sum_{j=1}^{N_1} d_j x_{l-j} + \sum_{j=1}^{N_2} h_j x_{l+s+j}, \quad (9)$$

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

$$M \left( x_{l+m} - \sum_{j=1}^{N_1} d_j x_{l-j} - \sum_{j=1}^{N_2} h_j x_{l+s+j} \right)^2 \rightarrow \min_{\{d_j, h_j\}}. \quad (10)$$

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

$$\begin{aligned} \sum_{j=1}^{N_1} d_j R_r(j-k) + \sum_{j=1}^{N_2} h_j R_r(s+j+k) &= R_r(m+k), 1 \leq k \leq N_1 \\ \sum_{j=1}^{N_1} d_j R_r(s+j+k) + \sum_{j=1}^{N_2} h_j R_r(j-k) &= R_r(m-s-k), 1 \leq k \leq N_2, \end{aligned} \quad (11)$$

решение которой не предоставляет труда. Если коэффициенты  $\{a_k(\tau), 1 \leq k \leq n, 1 \leq \tau \leq T\}$  и дисперсии  $\{\sigma_\tau^2, 1 \leq \tau \leq T\}$  известны, то ковариации  $R_r(k)$  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 ПК и их работоспособность доказана методами имитационного моделирования.

## ЛИТЕРАТУРА

1. Меленец, Ю. В. Представление квазипериодических случайных процессов уравнениями авторегрессии / Ю. В. Меленец / Белорусский государственный университет. Мн., 1982. 13 с. Деп. в ВИНТИ 25.11.82 №5833-82.
2. Pagano, M. J. On periodic and multiple autoregressions / M. J. Pagano // The Annals of Statistics. 1988. V.6. № 1. P. 1310–1317.
3. Troutman B. M. Some results in periodic autoregressions / B. M. Troutman // Biometrika. 1989. V. 66. N. 2. P. 219-228.
4. Андерсон, Т. Статистический анализ временных рядов / Т. Андерсон. М.: Мир, 1998. 512 с.