УПРАВЛЯЮЩИЕ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ

УДК 519.246.2

НЕАСИМПТОТИЧЕСКОЕ ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ АВТОРЕГРЕССИОННОГО ПРОЦЕССА В СЛУЧАЕ НЕИЗВЕСТНОЙ ДИСПЕРСИИ ШУМА

© С. Э. Воробейчиков, А. В. Пупков

Томский государственный университет,
634050, г. Томск, просп. Ленина, 36
E-mail: sev@mail.tsu.ru, andrewpupkov@gmail.com

Рассматривается неасимптотическая процедура построения доверительной области параметра гауссовского процесса авторегрессии p-го порядка AR(p) в случае неизвестной дисперсии шума процесса. Процедура доверительного оценивания базируется на мартингальном свойстве числителя уклонения оценки метода наименьших квадратов. Приведены результаты численного моделирования.

Ключевые слова: авторегрессия, оценка МНК, последовательное оценивание, неасимптотическая доверительная область.

DOI: 10.15372/AUT20230405

Введение. Часто в инженерных приложениях приходится прибегать к аппроксимации случайных процессов моделями, линейно зависящими от параметров. Типичными представителями такого класса стохастических процессов выступают временны́е ряды авторегрессионного типа. При решении задачи аппроксимации ключевым является нахождение оценок параметров авторегрессии при заданном уровне априорного знания о процессе. Существует множество качественных методов оценивания параметров авторегрессии, к которым можно отнести метод максимального правдоподобия (ММП), метод наименьших квадратов (МНК), метод стохастической аппроксимации. Следует отметить, что асимптотическая теория оценивания авторегрессионных параметров разработана достаточно полно [1], но при конечном объёме выборки качество оценивания является неразрешённой проблемой.

В последнее время были предложены различные подходы для доверительного оценивания параметров авторегрессии [2–4] при различных предположениях о распределении процесса и структуре авторегрессионной схемы. В частности, в работе [2] рассматривается задача доверительного оценивания параметра авторегрессии первого порядка в случае положительно-определённого шума. Асимптотически точный доверительный интервал получен на основе оценки экстремальных значений (extreme value estimate) параметра авторегрессии. Отметим, что в качестве распределения шума в [2] рассматривается широкий класс моделей, параметры которого оцениваются путём применения ММП для порядковых статистик. В [3] рассматривается задача построения унифицированной асимптотически точной доверительной области параметров процесса авторегрессии первого порядка, распределение которой не зависит от наличия свободного члена и положения корреляционного параметра на числовой прямой в авторегрессионной схеме. Для решения используется эмпирический метод правдоподобия. В [4] предложена процедура построения доверительных интервалов для параметров авторегресии и получено асимтотическое распределение в виде $\exp(-\mathbf{e}^{-x})$.

Важным направлением в области исследований временны́х рядов выступает так называемое последовательное оценивание, позволяющее решить задачу идентификации параметров линейных систем в неасимптотической постановке. При использовании последовательных процедур оценивания подразумевается, что объём выборки не фиксирован и определяется специальным правилом остановки. Последовательные методы оценивания показали своё превосходство над классическими оценками (МНК, ММП), строящимися на основе фиксированного объёма выборки, поскольку в общем случае первые отличаются от последних ограниченностью среднеквадратической ошибки функцией вида $1/h$, где h — параметр, определяющий точность оценивания, увеличение которого приводит к росту среднего объёма выборки. К настоящему времени разработаны последовательные процедуры точечного и доверительного оценивания параметров линейных систем с заданной среднеквадратической точностью [5–12].

Отметим, что авторегрессионные схемы являются достаточно адекватными для описания поведения реальных объектов. В работе [13] предложена процедура, позволяющая выбрать наиболее подходящую модель для описания реального объекта. Процедура основана на минимизации расстояния между оценкой параметра авторегрессионной схемы и областью допустимых значений (ОДЗ).

Одной из проблем при построении оценок параметров является отсутствие априорного знания об уровне дисперсии шума. В работе [14] приведён алгоритм оценивания дисперсии шума, используемый в процедуре фильтрации сигнала, основанный на дискретном преобразовании Фурье. В [15] предложена итерационная процедура оценивания дисперсии аддитивной ошибки зашумлённой авторегрессии.

Цель данной работы — представить процедуру, позволяющую решить задачу неасимптотического доверительного оценивания параметра гауссовской авторегрессии p-го порядка в случае неизвестной дисперсии шума.

Постановка задачи. Рассмотрим процесс авторегрессии p-го порядка:

$$x_k = X'_{k-1}\theta + b\varepsilon_k, \qquad k \geqslant 1, \tag{1}$$

где $X_k = (x_k, \ldots, x_{k-p+1})'$; $\theta = (\theta_1, \ldots, \theta_p)'$ — вектор неизвестных параметров (штрих ($'$) означает транспонирование); b^2 — неизвестная дисперсия шума процесса; $\{\varepsilon_k\}$ — последовательность независимых стандартных гауссовских случайных величин, не зависящая от начальных значений процесса $x_0, x_{-1}, \ldots, x_{1-p}$. Ставится задача построения доверительной области для параметра θ в неасимптотической постановке.

Оценивание дисперсии шума процесса. На первом этапе введём пробную оценку параметра θ на основе фиксированного объёма выборки n_1 по методу наименьших квадратов:

$$\tilde{\theta}_{n_1} = \Big(\sum_{k=1}^{n_1} X_{k-1}X'_{k-1}\Big)^{-1} \sum_{k=1}^{n_1} X_{k-1}x_k, \tag{2}$$

и используем её для оценивания дисперсии шума процесса. На втором этапе построим оценку дисперсии по схеме

$$\Gamma_{n_1,n_2} = \frac{S_{n_1,n_2}}{n_2-2}, \qquad S_{n_1,n_2} = \sum_{k=n_1+1}^{n_1+n_2} (x_k - X'_{k-1}\tilde{\theta}_{n_1})^2. \tag{3}$$

По аналогии с результатом Теоремы 1 в [7] получим верхнюю границу для функции распределения величины S_{n_1,n_2}.

Лемма 1. Пусть случайная величина S_{n_1,n_2} определена выражением (3), тогда её функция распределения имеет оценку сверху вида

$$\mathsf{P}\,(S_{n_1,n_2} < z) \leqslant \mathsf{P}\Big(b^2 \sum_{k=n_1+1}^{n_1+n_2} \varepsilon_k^2 < z\Big).$$

Доказательство. Пусть $N = n_1+n_2$, $D = \max(0, z-S_{n_1,n_2-1})$, $f_{\varepsilon_t|T}(x)$ — условная плотность распределения величины ε_t при условии $\mathcal{F}_T = \sigma\{X_0, \varepsilon_1, \ldots, \varepsilon_T\}$. Учитывая унимодальность и симметричность плотности $f_{\varepsilon_t|T}(x)$, $T < t$, имеем

$$\mathsf{P}\Big(\sum_{k=n_1+1}^{n_1+n_2} (x_k - X'_{k-1}\tilde{\theta}_{n_1})^2 < z\Big) = \mathbb{E}\,\mathsf{P}\,((x_N - X'_{N-1}\tilde{\theta}_{n_1})^2 < z - S_{n_1,n_2-1} \mid \mathcal{F}_{N-1}) =$$

$$= \mathbb{E}\,\mathsf{P}\,(-\sqrt{D} < X'_{N-1}(\theta - \tilde{\theta}_{n_1}) + b\varepsilon_N < \sqrt{D} \mid \mathcal{F}_{N-1}) =$$

$$= \mathbb{E} \int_{-\sqrt{D}}^{\sqrt{D}} f_{\varepsilon_N|N-1}(X'_{N-1}(\theta - \tilde{\theta}_{n_1}) + bx)\,dx \leqslant$$

$$\leqslant \mathbb{E} \int_{-\sqrt{D}}^{\sqrt{D}} f_{\varepsilon_N|N-1}(x)\,dx = \mathbb{E}\,\mathsf{P}\,(-\sqrt{D} < b\varepsilon_N < \sqrt{D} \mid \mathcal{F}_{N-1}) = \mathsf{P}\,(b^2\varepsilon_N^2 + S_{n_1,n_2-1} < z).$$

Продолжая аналогичные рассуждения, приходим к результату Леммы 1.

Последовательная оценка МНК. Далее, используя оценку (3), получим набор моментов остановки по схеме

$$\tau_0(h) = n_1 + n_2, \quad \tau_l(h) = \inf\Big\{n > \tau_{l-1}(h) : \sum_{k=\tau_{l-1}(h)+1}^{n} \frac{x_{k-l}^2}{\Gamma_{n_1,n_2}} \geqslant h\Big\}, \quad l = 1, \ldots, p, \tag{4}$$

для любого $h > 0$. Затем на каждом из интервалов $[(\tau_{l-1}(h), \tau_l(h)]$, $l = 1, \ldots, p$, введём систему коэффициентов

$$\beta_k^{(l)}(h) = \begin{cases} 1, & \text{если } \tau_{l-1}(h) < k < \tau_l(h); \\ \alpha_l(h), & \text{если } k = \tau_l(h), \end{cases} \tag{5}$$

где поправочный коэффициент $0 < \alpha_l(h) \leqslant 1$ единственным образом находится из уравнения

$$\sum_{k=\tau_{l-1}(h)+1}^{\tau_l(h)-1} \frac{x_{k-l}^2}{\Gamma_{n_1,n_2}} + \alpha_l(h)\frac{x_{\tau_l(h)-l}^2}{\Gamma_{n_1,n_2}} = h, \qquad l = 1, \ldots, p.$$

Для дальнейших выводов введём обозначения. Пусть $\langle a\rangle_l$ — l-я координата вектора a, а $\langle A\rangle_{l,s}$ — элемент l-й строки и s-го столбца матрицы A.

Последовательную модификацию оценки МНК представим в виде

$$\hat{\theta}(h) = G^{-1}(h)\vartheta(h), \tag{6}$$

где l-я координата вектора $\vartheta(h)$ определяется формулой

$$\langle\vartheta(h)\rangle_l = \sum_{k=\tau_{l-1}(h)+1}^{\tau_l(h)} \sqrt{\beta_k^{(l)}(h)}\,\frac{x_{k-l}x_k}{\sqrt{\Gamma_{n_1,n_2}}}, \qquad l=1,\ldots,p, \tag{7}$$

а элементы матрицы $G(h)$ вычисляются следующим образом:

$$\langle G(h)\rangle_{l,s} = \sum_{k=\tau_{l-1}(h)+1}^{\tau_l(h)} \sqrt{\beta_k^{(l)}(h)}\,\frac{x_{k-l}x_{k-s}}{\sqrt{\Gamma_{n_1,n_2}}}, \qquad l=1,\ldots,p, \quad s=1,\ldots,p.$$

Подставив уравнение процесса (1) в (7), получим разложение

$$\langle\vartheta(h)\rangle_l = \sum_{k=\tau_{l-1}(h)+1}^{\tau_l(h)} \sqrt{\frac{\beta_k^{(l)}(h)}{\Gamma_{n_1,n_2}}}\,x_{k-l}\sum_{s=1}^{p} x_{k-s}\theta_s + b\sum_{k=\tau_{l-1}(h)+1}^{\tau_l(h)} \sqrt{\frac{\beta_k^{(l)}(h)}{\Gamma_{n_1,n_2}}}\,x_{k-l}\varepsilon_k =$$

$$= \sum_{s=1}^{p}\theta_s\langle G(h)\rangle_{l,s} + b\langle\zeta(h)\rangle_l, \tag{8}$$

$$\langle\zeta(h)\rangle_l = \sum_{k=\tau_{l-1}(h)+1}^{\tau_l(h)} \sqrt{\frac{\beta_k^{(l)}(h)}{\Gamma_{n_1,n_2}}}\,x_{k-l}\varepsilon_k, \qquad l=1,\ldots,p.$$

Запишем разложение (8) в матричном виде:

$$\vartheta(h) = G(h)\theta + b\zeta(h). \tag{9}$$

Выделив $\vartheta(h)$ из оценки (6) и подставив в (9), имеем

$$G(h)(\hat\theta(h) - \theta) = b\zeta(h). \tag{10}$$

Результат следующей леммы является основой для построения неасимптотической доверительной области параметра θ.

Лемма 2. Пусть в процессе (1) последовательность $\{\varepsilon_k\}$ — это независимые стандартные гауссовские величины. Тогда для любого $h > 0$ случайный вектор $h^{-1/2}\zeta(h)$ имеет стандартное гауссовское распределение, т. е. $\mathrm{Law}\,(h^{-1/2}\zeta(h)) \sim \mathcal{N}(0, I_p)$, где I_p — единичная матрица размера p.

Доказательство. Введём набор процессов $\{M_l(n)\}_{n>n_1+n_2}$, $l=1,\ldots,p$, где

$$M_l(n) = \sum_{k=n_1+n_2+1}^{n} \frac{x_{k-l}\varepsilon_k}{\sqrt{\Gamma_{n_1,n_2}}}.$$

Покажем, что процессы $\{M_l(n), \mathcal{F}_n\}$ являются квадратично-интегрируемыми мартингалами. Для этого достаточно показать, что приращения $\Delta M_l(n) = M_l(n) - M_l(n-1)$ имеют нулевое условное математическое ожидание и конечную условную дисперсию. Введём поток σ-алгебр $\{\mathcal{F}_n\} = \sigma\{X_0, \varepsilon_1, \ldots, \varepsilon_n\}$, тогда

$$\mathbb{E}(\Delta M_l(n) \mid \mathcal{F}_{n-1}) = \Gamma_{n_1,n_2}^{-1/2} x_{n-l} \mathbb{E}(\varepsilon_n \mid \mathcal{F}_{n-1}) = 0,$$

$$\mathbb{E}([\Delta M_l(n)]^2 \mid \mathcal{F}_{n-1}) = \Gamma_{n_1,n_2}^{-1} x_{n-l}^2 \mathbb{E}(\varepsilon_n^2 \mid \mathcal{F}_{n-1}) = \Gamma_{n_1,n_2}^{-1} x_{n-l}^2.$$

Далее заметим, что случайный вектор $h^{-1/2}\zeta(h)$, который формируется на основе процессов $\{M_l(n)\}$, по построению удовлетворяет условиям Теоремы 3 в [8]. Данный факт доказывает результат Леммы 2.

Построение доверительной области. Далее получим доверительную область для параметра θ. Рассмотрим статистику вида

$$Z(h) = \frac{1}{h\sqrt{\Gamma_{n_1,n_2}}} \|G(h)(\hat{\theta}(h) - \theta)\|_2, \tag{11}$$

где $\|a\|_2$ — евклидова норма вектора a, определяемая выражением

$$\|a\|_2 = \Big(\sum_l \langle a\rangle_l^2\Big)^{1/2}.$$

Найдём нижнюю границу функции распределения статистики (11). Используя матричное уравнение (10), приходим к равенствам

$$\mathsf{P}\left(Z(h) < z\right) = \mathsf{P}\Big(\frac{1}{h\sqrt{\Gamma_{n_1,n_2}}} \|G(h)(\hat{\theta}(h) - \theta)\|_2 < z\Big) =$$

$$= \mathsf{P}\Big(\frac{b\|\zeta(h)\|_2}{h\sqrt{\Gamma_{n_1,n_2}}} < z\Big) = \mathbb{E}\,\mathsf{P}\Big(\frac{\|\zeta(h)\|_2}{\sqrt{h}} < z\sqrt{\frac{h\Gamma_{n_1,n_2}}{b^2}} \;\Big|\; \mathcal{F}_{n_1+n_2}\Big).$$

Далее пусть $G(y)$ — функция распределения величины $\Gamma_{n_1,n_2}/b^2$, тогда, используя результат Леммы 1, получим цепочку неравенств:

$$\mathbb{E}\,\mathsf{P}\Big(\frac{\|\zeta(h)\|_2}{\sqrt{h}} < z\sqrt{\frac{h\Gamma_{n_1,n_2}}{b^2}} \;\Big|\; \mathcal{F}_{n_1+n_2}\Big) = \int_0^\infty \mathsf{P}\Big(\frac{\|\zeta(h)\|_2}{\sqrt{h}} < z\sqrt{hy} \;\Big|\; \mathcal{F}_{n_1+n_2}\Big)\, dG(y) \geqslant$$

$$\geqslant \int_0^\infty \mathsf{P}\Big(\frac{\|\zeta(h)\|_2}{\sqrt{h}} < z\sqrt{hy} \;\Big|\; \mathcal{F}_{n_1+n_2}\Big)\, d\,\mathsf{P}\Big(\sum_{k=n_1+1}^{n_1+n_2} \varepsilon_k^2 < y(n_2-2)\Big).$$

Учитывая тот факт, что случайная величина $h^{-1/2}\|\zeta(h)\|_2$ имеет χ-распределение с p степенями свободы в соответствии с результатом Леммы 2, а $\sum\limits_{k=n_1+1}^{n_1+n_2} \varepsilon_k^2$ имеет χ^2-распределение с n_2 степенями свободы, то получаем основной результат в виде следующей теоремы.

Теорема 1. Пусть $\{\varepsilon_k\}$ — последовательность независимых стандартных гауссовских случайных величин. Тогда для последовательного плана оценивания параметра авторегресии, определяемого совокупностью моментов остановки (4), системой коэффициентов (5) и точечной оценкой (6), для любого $h > 0$ и любого $z > 0$ выполняется неравенство

$$\inf_{\theta \in \Theta^p} \mathsf{P}\Big(\frac{1}{h\sqrt{\Gamma_{n_1,n_2}}}\,\|G(h)(\hat{\theta}(h) - \theta)\|_2 < z\Big) \geqslant q(z, h; n_2, p);$$

$$q(z, h; n_2, p) = \frac{1}{2^{n_2/2}\Gamma(n_2/2)} \int\limits_0^\infty \Psi_p\Big(z\sqrt{\frac{yh}{n_2 - 2}}\,\Big) y^{n_2/2-1} \mathbf{e}^{-y/2}\, dy, \tag{12}$$

где $\Theta^p \subset \mathbb{R}^p$ — некоторая подобласть евклидова пространства; n_2 — объём выборки для оценивания дисперсии шума; $\Gamma(x)$ — гамма-функция; $\Psi_p(x)$ — функция распределения, определяемая выражением

$$\Psi_p(x) = \frac{1}{2^{p/2-1}\Gamma(p/2)} \int\limits_0^x y^{p-1} \mathbf{e}^{-y^2/2}\, dy. \tag{13}$$

З а м е ч а н и е. Отметим, что при условии стационарности процесса (1) отношение $G(h)/h$ сходится к невырожденной матрице почти наверное в соответствии с результатом Предложения 2 в [9].

Численные результаты. Для представления численных результатов введём следующие обозначения. Пусть в качестве p_1 выступает значение нижней границы доверительной вероятности $q(z, h; n_2, p)$, определяемой выражением (12). Величина $\bar{Z}(h)$ — усреднение базовой статистики (11). Усреднённое значение последнего момента остановки $\tau_p(h)$ процедуры, определяемое правилом (4), обозначим через $\bar{\tau}_p$. Величина $\tau_p(h)$ задаёт объём необходимой выборки для построения доверительного эллипсоида. Величина Δ задаёт выборочную норму уклонения оценки $\|\hat{\theta}(h) - \theta\|_2$, а величина $\bar{p}_z$ — выборочную вероятность накрытия параметра θ доверительным p-мерным эллипсоидом при фиксированном значении параметра z, который отвечает за длину полуосей эллипсоида. Если переформулировать, то величина $\bar{p}_z$ показывает выборочную вероятность события $\{Z(h) < z\}$. Величина $\bar{\mu}_{\max}$ описывает усреднённое значение максимальной полуоси эллипсоида вида

$$\|G(h)(\hat{\theta}(h) - \theta)\|_2^2 = z^2 h^2 \Gamma_{n_1,n_2}.$$

Длины полуосей эллипсоида μ_i, $i = 1, \ldots, p$, определяются по формуле

$$\mu_i = hz\sqrt{\frac{\Gamma_{n_1,n_2}}{\lambda_i}},$$

где λ_i — соответствующие собственные значения матрицы $W = G'(h)G(h)$.

В табл. 1 представлены выборочные характеристики процедуры (12) для процесса AR(3) вида

$$x_k = 0,6x_{k-1} - 0,8x_{k-2} + 0,5x_{k-3} + b\varepsilon_k, \qquad k \geqslant 1, \tag{14}$$

Таблица 1

Усреднённые характеристики доверительной области (12) для процесса (14)

h	p_1	$b = 0{,}8$					$b = 2{,}4$				
		$\bar{Z}(h)$	$\bar{\tau}_p$	Δ	$\bar{p}_z$	$\bar{\mu}_{\max}$	$\bar{Z}(h)$	$\bar{\tau}_p$	Δ	$\bar{p}_z$	$\bar{\mu}_{\max}$
100	0,039	0,140	218,5	0,256	0,062	0,164	0,144	218,5	0,277	0,055	0,169
500	0,299	0,065	914,2	0,120	0,355	0,165	0,063	935,7	0,117	0,351	0,166
1000	0,564	0,046	1771,1	0,084	0,640	0,166	0,044	1772,0	0,081	0,661	0,166
1500	0,730	0,039	2723,2	0,071	0,768	0,166	0,037	2713,3	0,069	0,786	0,166
2000	0,831	0,032	3637,5	0,059	0,868	0,165	0,031	3554,7	0,056	0,887	0,165
2500	0,892	0,028	4518,6	0,052	0,934	0,166	0,029	4421,5	0,055	0,908	0,166
3000	0,930	0,026	5454,4	0,049	0,954	0,165	0,026	5452,4	0,048	0,954	0,166
3500	0,953	0,024	6280,6	0,045	0,959	0,166	0,024	6366,4	0,046	0,968	0,165

Таблица 2

Усреднённые характеристики доверительной области (12) для процесса AR(3)

θ_1	θ_2	θ_3	Неизвестная дисперсия b^2					Известная дисперсия b^2				
			$\bar{Z}(h)$	$\bar{\tau}_p$	Δ	$\bar{p}_z$	$\bar{\mu}_{\max}$	$\bar{Z}(h)$	$\bar{\tau}_p$	Δ	$\bar{p}_z$	$\bar{\mu}_{\max}$
–0,7	–0,7	–0,9	0,047	742,8	0,067	0,965	0,235	0,055	434,1	0,080	0,961	0,245
–0,6	–0,7	–0,8	0,047	956,4	0,075	0,971	0,269	0,054	626,3	0,084	0,963	0,268
–0,5	–0,6	–0,8	0,047	1058,6	0,067	0,970	0,232	0,055	686,2	0,078	0,967	0,230
–0,4	–0,3	–0,8	0,048	1196,4	0,073	0,961	0,231	0,055	758,7	0,087	0,955	0,240
–0,3	–0,6	–0,5	0,048	1837,6	0,074	0,970	0,251	0,053	1268,6	0,080	0,970	0,250
–0,1	0	–0,8	0,048	1273,1	0,064	0,964	0,179	0,055	803,5	0,074	0,970	0,183
0,1	–0,3	0,7	0,048	1599,3	0,063	0,968	0,203	0,055	1060,0	0,071	0,962	0,203
0,3	0,5	–0,7	0,048	1412,1	0,075	0,975	0,253	0,054	885,8	0,084	0,965	0,254
0,4	0,5	–0,8	0,046	1138,9	0,062	0,971	0,206	0,055	739,1	0,073	0,969	0,208
0,5	0,3	–0,6	0,049	1858,5	0,077	0,965	0,248	0,055	1273,8	0,087	0,962	0,246
0,6	–0,6	0,9	0,047	739,0	0,062	0,971	0,204	0,054	459,4	0,071	0,955	0,210
0,9	–0,9	0,9	0,047	622,9	0,081	0,969	0,304	0,055	384,6	0,098	0,967	0,323

для различных значений дисперсии шума b^2 и различных значений параметра h ($z = 0{,}05$, $n_1 = 20$, $n_2 = 15$), определяющего длительность процедуры (12). Отметим, что при увеличении h возрастает нижняя граница доверительной вероятности p_1, а также становятся более стабильными границы эллипсоида в силу Замечания. Значения величины $\bar{p}_z$ незначительно больше, чем соответствующие значения нижней границы доверительной вероятности p_1, что свидетельствует о близости нижней границы доверительной вероятности p_1 и истинного её значения. При различных значениях дисперсии $b^2 \in \{0,8^2, 2,4^2\}$ объём выборки, необходимый для доверительного оценивания, аналогичен в силу нормировки наблюдений на оценку дисперсии (3) при построении моментов остановки (4).

В табл. 2 представлены результаты моделирования при построении доверительного эллипсоида для процесса авторегрессии третьего порядка AR(3) при различных значениях параметра θ при условии неизвестной и известной дисперсий шума b^2 ($h = 855$, $z = 0{,}1$, $p_1 = 0{,}95$, $b = 2{,}4$, $n_1 = 20$, $n_2 = 15$). Совокупность первых трёх столбцов показывает комбинацию значений параметров θ_1, θ_2 и θ_3. Совокупность следующих пяти столбцов демонстрирует результаты моделирования в случае, когда дисперсия шума неизвестна, т. е.

Таблица 3

Усреднённые характеристики доверительной области (12) для процесса AR(2)

θ_1	θ_2	$p = 2$, $p_1 = 0{,}977$					$p = 3$, $p_1 = 0{,}950$				
		$\bar{Z}(h)$	$\bar{\tau}_p$	Δ	$\bar{p}_z$	$\bar{\mu}_{\max}$	$\bar{Z}(h)$	$\bar{\tau}_p$	Δ	$\bar{p}_z$	$\bar{\mu}_{\max}$
–1,8	–0,9	0,039	96,2	0,891	0,977	3,590	0,047	140,3	10,514	0,976	91,364
–1,4	–0,6	0,039	400,7	0,196	0,986	0,793	0,048	618,5	1,626	0,971	4,491
–1	–0,2	0,040	719,5	0,156	0,978	0,596	0,049	1130,8	0,286	0,972	1,078
–0,5	0,3	0,040	1093,8	0,093	0,978	0,350	0,050	1715,1	0,137	0,97	0,387
0	0,8	0,039	968,1	0,039	0,983	0,109	0,048	1516,9	0,132	0,963	0,505
0,7	–0,3	0,040	1495,1	0,060	0,988	0,216	0,050	2340,9	0,099	0,968	0,363
1,5	–0,7	0,039	308,3	0,203	0,987	0,823	0,047	482,4	5,043	0,978	12,428
1,8	–0,9	0,037	98,9	0,710	0,983	3,220	0,046	138,8	8,663	0,971	51,302

в процедуре используется оценка дисперсии (3). Совокупность оставшихся пяти столбцов включает в себя результаты, когда известна априорная информация о дисперсии шума процесса, т. е. оценка дисперсии шума Γ_{n_1,n_2} в процедуре в точности равна истинному значению оцениваемой дисперсии b^2. При моделировании процедуры в случае известной дисперсии первые n_1+n_2 элемента выборки не использовались при построении последовательной оценки, но учитывались при расчёте необходимого объёма выборки $\tau_p(h)$. Легко заметить, что при увеличении априорной информации уменьшается требуемый объём выборки для получения доверительной области с аналогичными свойствами. Данный факт является вполне естественным, поскольку величина S_{n_1,n_2} принимает своё минимальное значение при условии, что пилотная оценка, определяемая выражением (2), равна истинному значению θ. Следовательно, расстояние между моментами $\{\tau_l\}$ будет тем меньше, чем лучше будет оценён параметр θ на первом шаге. Параметры h и z подобраны таким образом, чтобы нижняя граница доверительной вероятности была равна $p_1 = 0,95$. Добавим, что качество оценивания заметно ухудшается (в смысле увеличения длины полуосей эллипсоида) при приближении значений параметра θ к границам устойчивости процесса (1).

В табл. 3 продемонстрированы усреднённые характеристики доверительной области (12) для процесса авторегрессии при различных значениях параметра θ в условиях различных априорных предположений о порядке модели ($h = 855$, $z = 0{,}1$, $b = 2{,}4$, $n_1 = 20$, $n_2 = 15$). Выборка строилась по схеме процесса авторегрессии второго порядка, заданного уравнением

$$x_k = \theta_1 x_{k-1} + \theta_2 x_{k-2} + 2,4\varepsilon_k, \qquad k \geqslant 1.$$

Комбинации значений параметров θ_1 и θ_2 представлены в первых двух столбцах таблицы. В следующих пяти столбцах находятся усреднённые характеристики доверительного оценивания в случае, когда порядок авторегрессии оценён верно. В следующих пяти столбцах находятся характеристики доверительного оценивания параметров в случае, когда предположение о порядке авторегрессии ошибочно, т. е. фактически рассматривается ситуация, когда порядок процесса авторегрессии равен 3, но $\theta_3 \equiv 0$. Поскольку оценка доверительной вероятности зависит от значения порядка процесса p через функцию распределения (13), то это приводит к различным значениям нижней границы доверительной вероятности p_1. Сравнивая столбцы под заголовками $\bar{\mu}_{\max}$, приходим к выводу, что неверно подобранный порядок процесса p приводит к кратному увеличению длины полуосей доверительного эллипсоида.

Усреднение результатов проводилось по 1000 реализаций процедуры.

Заключение. В данной работе предложена последовательная процедура доверительного оценивания параметра процесса авторегресии p-го порядка в случае гауссовского шума с неизвестной дисперсией. Процедура позволяет построить эллипсоид, накрывающий неизвестный параметр θ с доверительной вероятностью в неасимптотической постановке. Нижняя граница доверительной вероятности рассчитывается с использованием трёх входных параметров процедуры: p — порядок авторегрессии, n_2 — объём выборки, необходимый для построения оценки дисперсии, h — параметр, отвечающий за объём выборки последовательной модификации оценки МНК параметра θ. Следовательно, при фиксированных значениях p и n_2 для построения доверительных границ параметра удовлетворительного качества необходимо подбирать параметры h и z. С одной стороны, чем больше значение параметра h, тем устойчивее границы доверительного эллипсоида. С другой стороны, сокращение значения z в (12) приводит к уменьшению длины максимальной полуоси эллипсоида, но одновременно с этим и к росту оценки доверительной вероятности, которую можно увеличивать за счёт повышения параметра h. Параметр z, отвечающий за длину полуосей доверительного эллипсоида, можно связать с нижней границей доверительной вероятности $q(z, h; n_2, p)$, определяемой формулой (12), при фиксированных значениях входных параметров $h = h_0$, $n_2 = n_2^0$ и $p = p_0$. Для этой цели достаточно найти значение обратной функции

$$z = q^{-1}(\rho, h_0; n_2^0, p_0),$$

где ρ — желаемая нижняя граница доверительной вероятности.

Полученные результаты могут быть использованы в приложениях теории оптимального управления, оценивания временны́х рядов и обнаружения разладки.

СПИСОК ЛИТЕРАТУРЫ

1. **Anderson T. W.** The Statistical Analysis of Time Series. New York: Wiley, 1971. 704 p.
2. **Hsiao W., Huang H., Ing C.** Interval estimation for a first-order positive autoregressive process // Journ. Time Series Analysis. 2018. **39**, Iss. 3. P. 447–467.
3. **Liu X., Peng L.** Asymptotic theory and unified confidence region for an autoregressive model // Journ. Time Series Analysis. 2019. **40**, Iss. 1. P. 43–65.
4. **Moritz J.** Simultaneous confidence bands for sequential autoregressive fitting // Journ. Multivariate Analysis. 2014. **124**. P. 130–149.
5. **Borisov V. Z., Konev V. V.** On sequential parameter estimation in discrete time processes // Automation and Remote Control. 1977. **38**, Iss. 10. P. 1475–1480.
6. **Конев В. В., Пергаменщиков С. М.** Последовательные планы идентификации параметров динамических систем // Автомат. и телемех. 1981. **7**, № 5. С. 84–92.
7. **Vorobeychikov S., Burkatovskaya Yu.** Non-asymptotic confidence estimation of the autoregressive parameter in AR(1) process with an unknown noise variance // Austrian Journ. Statistics. 2020. **49**, Iss. 4. P. 19–26.
8. **Konev V., Nazarenko B.** Sequential fixed accuracy estimation for nonstationary autoregressive processes // Annals of the Institute of Statist. Math. 2020. **72**, Iss. 1. P. 235–264.
9. **Konev V., Pupkov A.** Confidence estimation of autoregressive parameters based on noisy data // Automation and Remote Control. 2021. **82**, Iss. 6. P. 1030–1048.
10. **Konev V. V., Vorobeychikov S. E.** Fixed accuracy estimation of parameters in a threshold autoregressive model // Annals of the Institute of Statistical Mathematics. 2022. **74**. P. 685–711.

11. **Sriram T. N.** Sequential estimation of the autoregressive parameter in a first order autoregressive process // Sequential Analysis. 1988. **7**, Iss. 1. P. 53–74.
12. **Arkoun O.** Sequential adaptive estimators in nonparametric autoregressive models // Sequential Analysis. 2011. **30**, Iss. 2. P. 229–247.
13. **Тырсин А. Н., Серебрянский С. М.** Распознавание зависимостей во временных рядах на основе структурных разностных схем // Автометрия. 2015. **51**, № 2. С. 54–60.
14. **Воскобойников Ю. Е.** Оценивание оптимальных параметров пространственно-локальных алгоритмов фильтрации сигналов // Автометрия. 2019. **55**, № 3. С. 13–21. DOI: 10.15372/AUT20190302.
15. **Xia Y., Zheng W. X.** Novel parameter estimation of autoregressive signals in the presence of noise // Automatica. 2015. **62**. P. 98–105.

Поступила в редакцию 06.02.2023
После доработки 10.02.2023
Принята к публикации 14.03.2023