

СТАТИСТИЧЕСКИЙ АНАЛИЗ ОЦЕНОК СПЕКТРАЛЬНЫХ ПЛОТНОСТЕЙ ВРЕМЕННЫХ РЯДОВ

Н. В. Семенчук

Гродненский государственный университет

Гродно, Беларусь

E-mail: semenchuk_nata@mail.ru

Статья посвящена статистическому анализу оценок спектральных плотностей стационарных случайных процессов и исследованию возможностей их применения.

Ключевые слова: стационарный процесс, оценка спектральной плотности, окна просмотра данных, спектральные окна, процесс авторегрессии.

Построение оценок спектральных плотностей стационарных случайных процессов является одной из главных проблем в анализе временных рядов. Оцениванию спектральных плотностей стационарных случайных процессов посвящена многочисленная литература. Например, работы Т. Андерсона [1], Д. Бриллинджера, И. Г. Журбенко, Н. Н. Труша.

Пусть $X(0), X(1), \dots, X(T-1)$ – T последовательных, через равные промежутки времени, наблюдений за стационарным случайным процессом $X(t)$, $t \in \mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$ с $\mathbf{M}X(t) = 0$, $t \in \mathbf{Z}$ со спектральной плотностью $f(\lambda)$, $\lambda \in \Pi = [-\pi; \pi]$.

Статистика

$$I_T(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=0}^{T-1} X(t) e^{-i\lambda t} \right|^2, \lambda \in \Pi$$

называется периодограммой.

Из [2] известно, что

$$\mathbf{M}I_T(\lambda) = f(\lambda) + O(T^{-1}),$$

$$\mathbf{D}I_T(\lambda) = f^2(\lambda) + O(T^{-1}).$$

Таким образом, каким большим не было бы T , дисперсия будет стремиться к постоянному значению $f^2(\lambda)$. Поэтому невозможно получить оценку с меньшей дисперсией простым увеличением выборки.

Рассмотрим процесс авторегрессии порядка p :

$$\sum_{j=0}^p a_j X_{t-j} = \varepsilon_t, \quad (a_0 = 1), \quad (1)$$

где ε_t представляет собой последовательность независимых, одинаково распределенных случайных величин, $\varepsilon_t \sim N(0, \sigma^2)$.

Теоретическая спектральная плотность процесса AR(p) имеет вид:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=1}^p a_j e^{-i \cdot j \lambda} \right|^{-2}.$$

На рис. 1 представлена реализация процесса AR(1) длиной $T = 128$ с параметром $a_1 = 0,7$, $\varepsilon_t \sim N(0,1)$.

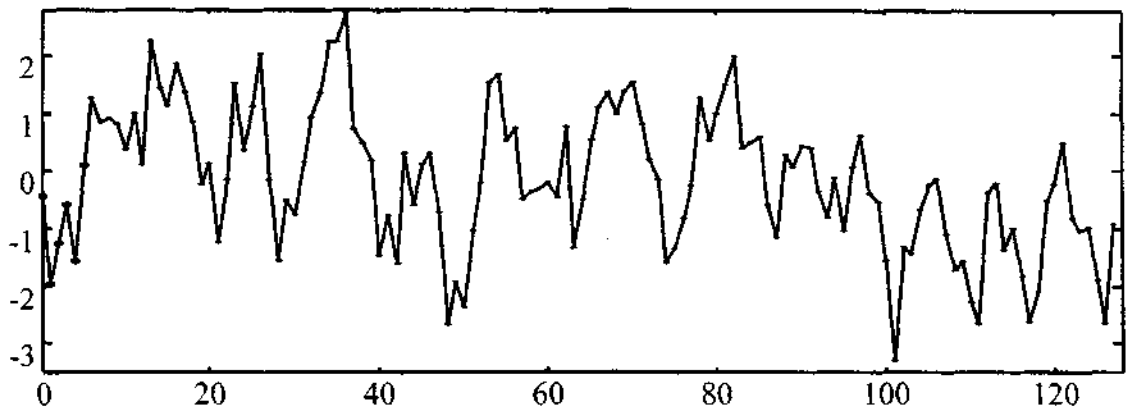


Рис. 1. Реализация процесса AR(1), $T = 128$

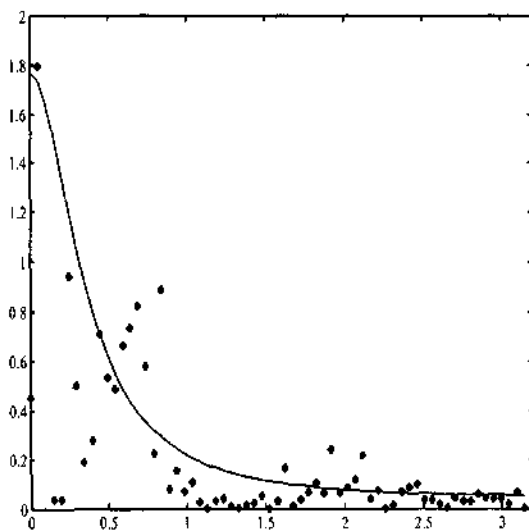


Рис. 2а. Теоретическая спектральная плотность и периодограмма процесса AR(1)

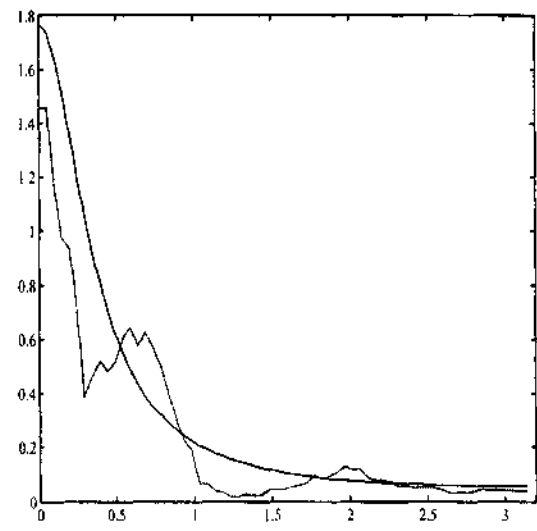


Рис. 2б. Теоретическая спектральная плотность и периодограмма сглаженная по 7 точкам процесса AR(1)

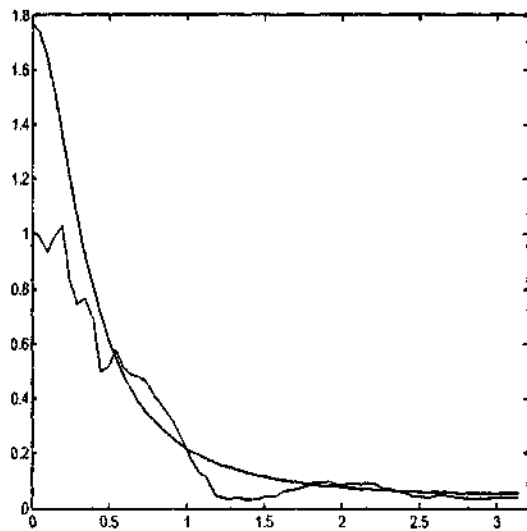


Рис. 2в. Теоретическая спектральная плотность и периодограмма сглаженная по 13 точкам процесса AR(1)

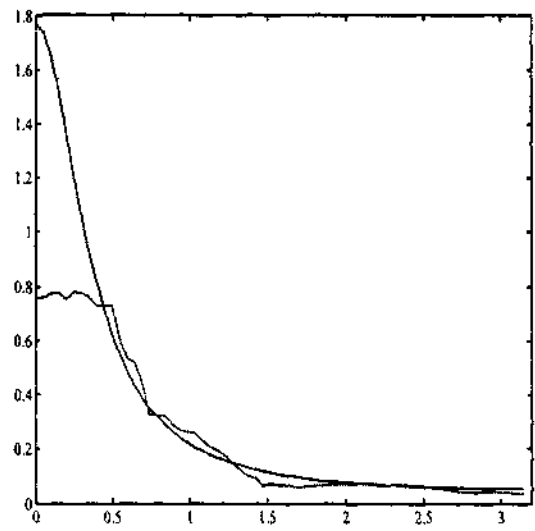


Рис. 2г. Теоретическая спектральная плотность и периодограмма сглаженная по 25 точкам процесса AR(1)

На рис. 2а представлена теоретическая спектральная плотность и значения периодограммы, построенной по реализации процесса с рис. 1 в точках $\lambda_j = \frac{2\pi j}{T}$.

Рассмотрим оценку спектральной плотности, построенную путем осреднения периодограммы спектральным окном. В качестве оценки спектральной плотности $f(\lambda)$ в точке $x = \lambda_j$ рассмотрим статистику вида

$$\hat{f}(\lambda_j) = \sum_{k=-M}^M \varphi^T(k) I_T(\lambda_{j+k}), \quad (2)$$

где $\varphi^T(k), k \in \mathbf{Z}$, – произвольная, не зависящая от наблюдений четная целочисленная функция, для $k \notin (-M, M)$, а $\lambda_j = \frac{2\pi j}{T}$, j – целое число, $-M \leq j \leq M$, M – целое число, $M \leq \left\lceil \frac{T}{2} \right\rceil$. Оценки такого вида рассматривались для одномерных стационарных процессов в работе И. Г. Журбенко [3] и Н. Н. Труша [4].

В общем случае $\varphi^T(k), k \in \mathbf{Z}$ строятся следующим образом:

$$\varphi_T(k) = \frac{\varphi\left(\frac{k}{M}\right)}{\sum_{j=-\infty}^{\infty} \varphi\left(\frac{j}{M}\right)},$$

где $\varphi(x)$ – четная, не зависящая от T действительная функция, для которой $\int_{-\infty}^{\infty} \varphi(x) dx = 1, k \leq M, M < \infty$.

Приведем некоторые примеры функций $\varphi(x)$, $x \in R$.

$$1. \varphi(x) = \begin{cases} \frac{1}{2}, & |x| \leq 1, \\ 0, & |x| > 1. \end{cases} \quad (3)$$

$$2. \varphi(x) = \begin{cases} 1 - |x|, & |x| < 1, \\ 0, & |x| \geq 1. \end{cases} \quad (4)$$

$$3. \varphi(x) = \begin{cases} \frac{3}{4}(1 - x^2), & |x| < 1, \\ 0, & |x| \geq 1. \end{cases} \quad (5)$$

$$4. \varphi(x) = \begin{cases} \frac{\pi}{4} \cos \frac{\pi x}{2}, & |x| < 1, \\ 0, & |x| \geq 1. \end{cases} \quad (6)$$

Осреднение $2M + 1$ соседних значений периодограмм с помощью окна (3) приводит к уменьшению дисперсии такой оценки в $2M + 1$ раз по сравнению с дисперсией периодограммы. Однако смещение оценки $f(\lambda)$ может возрастать с возрастанием параметра M , следовательно, в качестве значения M должно быть выбрано промежуточное значение.

Рис. 2б – 2г представляют спектральные оценки сглаженной периодограммы, приведенной на рис. 2а. Статистика вычислена при $M = 3, 6, 12$ для окна (3).

Если спектральная плотность меняется значительно, то лучше усреднять ординаты периодограммы с большим весом в непосредственной окрестности точки λ , чем на некотором расстоянии. Дисперсия оценки $f(\lambda)$ при больших выборках минимальна при использовании окна (3). Однако возможны ситуации, при которых оценка с окном (3) имеет большее смещение по сравнению с оценкой, использующей подходящий выбор окна (4) – (6).

На рис. 3 представлена реализация процесса AR(2) длиной $T = 128$ с параметрами $a_1 = 1, a_2 = -0,5$.

Для данного процесса были исследованы оценки спектральной плотности с использованием спектральных окон (3) – (6) при различных значениях M . В качестве критерия выбора лучшей оценки использовалась величина:

$$\left(f(\lambda_j) - Mf^{(T)}(\lambda_j) \right)^2 + D(f^{(T)}(\lambda_j)) \rightarrow \min. \quad (7)$$

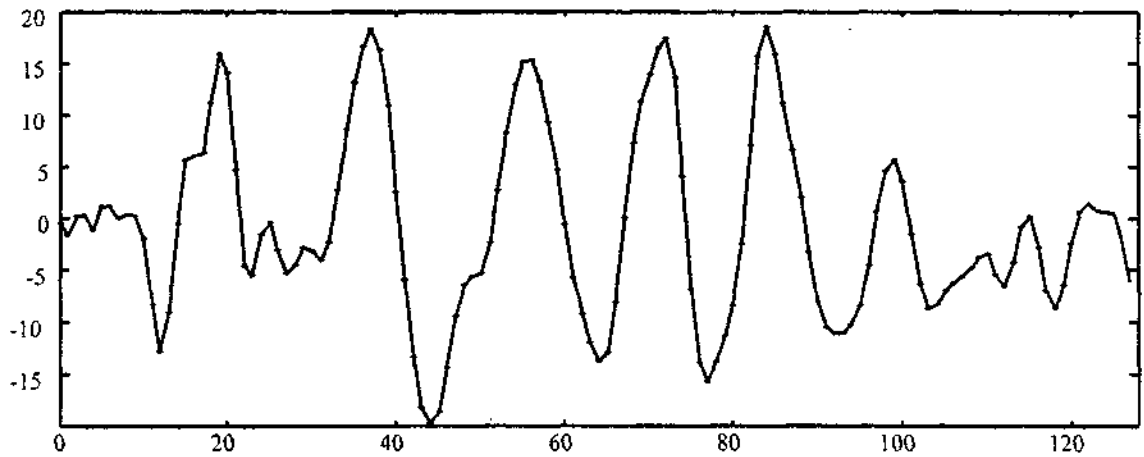


Рис. 3. Реализация процесса AR(2), $T = 128$

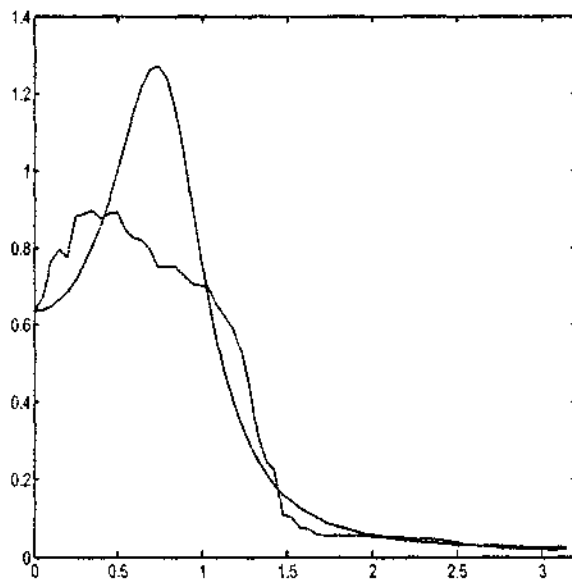


Рис. 4а. Теоретическая спектральная плотность и сглаженная окном (3) периодограмма процесса AR(2)

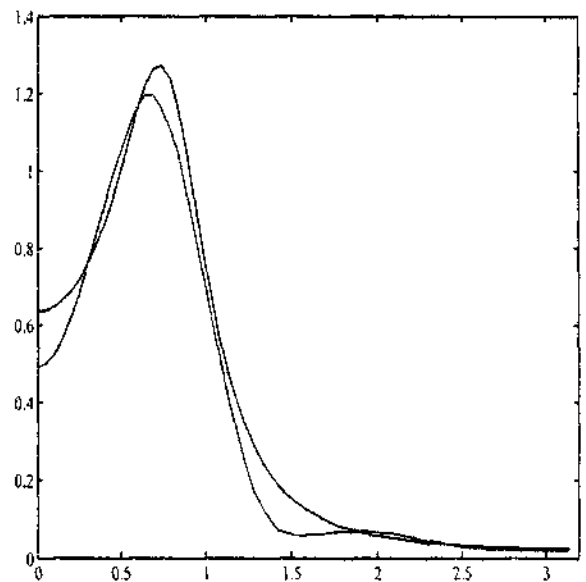


Рис. 4б. Теоретическая спектральная плотность и сглаженная окном (4) периодограмма процесса AR(2)

Для описанного процесса AR(2), по критерию (7), лучшей является оценка спектральной плотности, полученная сглаживанием периодограммы с помощью окна (4) при $M = 12$.

Рис. 4а, б представляют собой сглаженную окнами (3) и (4) соответственно периодограмму процесса AR(2).

Важно отметить, что вычисление $f(\lambda)$ для последовательности значений M дает гораздо большую информацию о спектре, чем вычисление при одном-единственном значении M . Графики оценок спектральной плотности при малых значениях M помогают раскрыть почти все периодические компоненты спектра и их местоположение, в то время как графики при больших значениях M дают гладкие кривые спектров и могут быть полезны при выборе модели.

Временное сглаживание наблюдаемых величин, предшествующее вычислению их преобразования Фурье, дает некоторые преимущества.

Определим расширенную периодограмму:

$$I_T^{(h)}(\lambda) = \frac{1}{2\pi H_2(0)} \left| \sum_{t=0}^{T-1} h_T(t) X(t) e^{-i\lambda t} \right|^2,$$

$$\lambda \in \Pi, H_2(0) = \sum_{t=0}^{T-1} h_T^2(t).$$

При определении расширенного конечного преобразования Фурье рассматривается функция $h_T(t) = h\left(\frac{t}{T}\right)$, называемая окном просмотра данных. Характерное поведение окна просмотра данных состоит в том, что функция $h(x)$ принимает максимальное значение, равное 1 в точке $x=0$, далее она монотонно убывает к 0 при $|x| \rightarrow 1$ и $h(x) = 0$ при $|x| > 1$.

Широкий спектр примеров окон просмотра данных приведен в [2].

Из [5] для модели (1) можем записать:

$$\sum_{j=0}^p a_j z^j = \prod_{j=1}^p (1 - u_j z), \quad u_j \in \mathbb{C}.$$

Если $u_j = \theta_j e^{-i\mu_j}$, то в этом случае спектральная плотность $f(\lambda)$ процесса (1) при $\theta_j \rightarrow 1$ имеет пики в частоте μ_j .

Спектральная плотность смоделированного процесса задается соотношением вида

$$f(\lambda) = \frac{\sigma^2}{2\pi} \prod_{j=1}^p \left| 1 - \theta_j e^{i(\lambda - \mu_j)} \right|^2.$$

На рис. 5 представлена реализация процесса AR(10) длиной $T = 128$, где $\varepsilon_t \sim N(0,1)$ и

$$u_1 = u_2 = 0.9e^{-i0.12\pi}, \quad u_3 = u_4 = \overline{u_1}$$

$$u_5 = u_6 = 0.9e^{-i0.4\pi}, \quad u_7 = u_8 = \overline{u_5}$$

$$u_9 = 0.9e^{-i0.8\pi}, \quad u_{10} = \overline{u_9}.$$

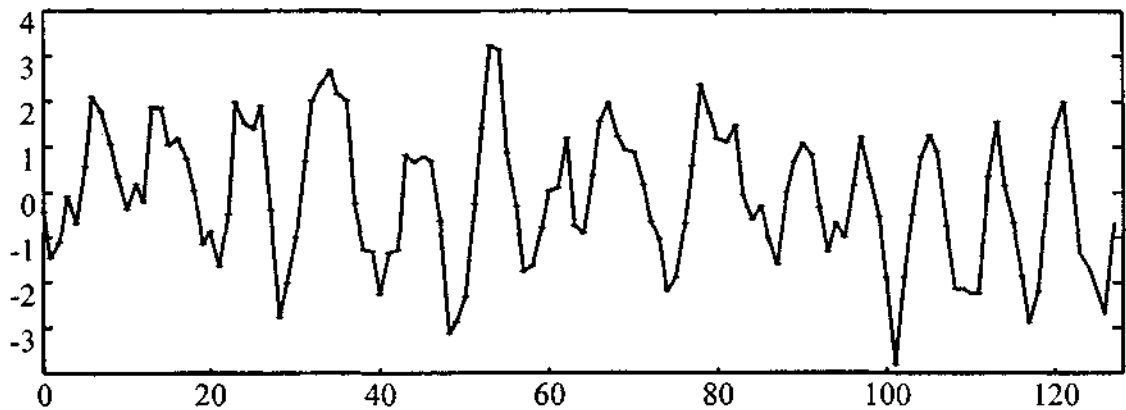


Рис. 5. Реализация процесса AR(10), $T = 128$

Исследовав оценки теоретической спектральной плотности процесса AR(10) для различных окон просмотра данных из [2, с. 62–63], а также для полиномиального окна [4, с. 133–136], получили, что лучшей по критерию (7) является оценка, построенная с полиномиальным окном просмотра данных.

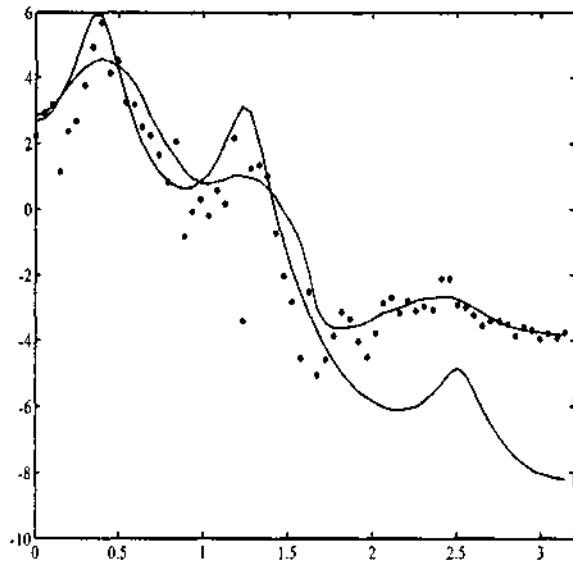


Рис. 6а. Теоретическая спектральная плотность процесса AR (10) и ее оценка без окон просмотра данных

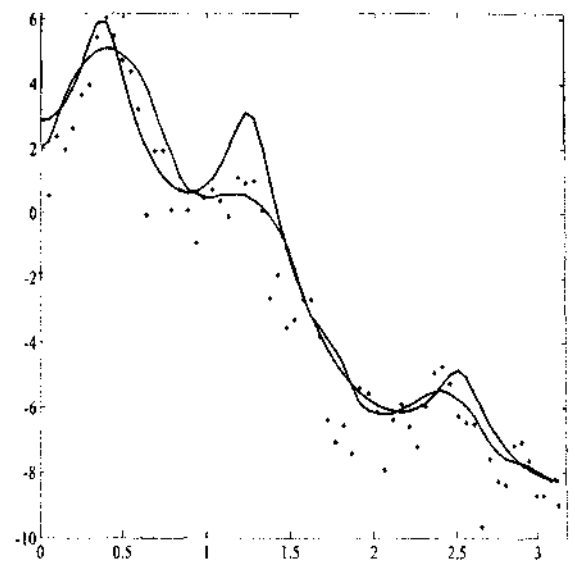


Рис. 6б. Теоретическая спектральная плотность процесса AR (10) и ее оценка с полиномиальным окном просмотра данных

Результат рис. 6а, б предложен для $\log I_T(\lambda)$ вместо $I_T(\lambda)$, так как преобразование логарифма делает дисперсию оценки асимптотически независимой от $f(\lambda)$. В данном случае график значений $\log f^{(T)}(\lambda)$ будет более показательным, чем график $f^{(T)}(\lambda)$.

Из рис. 6а, б видно, что использование полиномиального окна просмотра данных позволяет уменьшить смещение оценки, а также выявить скрытые пики спектральной плотности.

Для выявления скрытых особенностей спектральных плотностей можно также использовать оценки по непересекающимся и пересекающимся наблюдениям. Такие оценки для обычных периодограмм впервые предложил М. Барлетт. В [4] строится оценка спектральной плотности по непересекающимся и пересекающимся интервалам наблюдений с использованием произвольных окон просмотра данных.

Если число наблюдений T представимо в виде $T = LN$, где L – число интервалов, содержащих по N наблюдений (L не зависит от T), то в качестве оценки взаимной спектральной плотности $f(\lambda)$ будем рассматривать статистику

$$\hat{f}^{(T)}(\lambda) = \frac{1}{L} \sum_{l=0}^{L-1} I_T^{lN}(\lambda),$$

$\lambda \in \Pi$, которая построена путем осреднения расширенных периодограмм по L непересекающимся интервалам наблюдений. Аналогично вводится оценка, построенная по S пересекающимся интервалам наблюдений. Сглаживание наблюдений на каждом из интервалов наблюдения производится одним и тем же окном просмотра данных.

Оценки, построенные по непересекающимся и пересекающимся интервалам, являются асимптотически несмещенными, и их асимптотическая дисперсия соответственно в L и S раз меньше по сравнению с асимптотической дисперсией расширенной периодограммы, построенной по T наблюдениям. Данные оценки в случае рассмотренного выше процесса авторегрессии AR (10) дают похожий результат, т. е. позволяют выявлять скрытые пики спектральных плотностей случайных процессов.

ЛИТЕРАТУРА

1. *Андерсон Т.* Статистический анализ временных рядов. М.: Мир, 1976. 755 с.
2. *Бриллинджер Д.* Временные ряды. Обработка данных и теория. М.: Мир, 1980. 536 с.
3. *Журбенко И. Г.* Спектральный анализ временных рядов. М.: Изд-во МГУ, 1982. 168 с.
4. *Труш Н. Н.* Асимптотические методы статистического анализа временных рядов. Мн.: БГУ, 1999. 218 с.
5. *Dahlhaus R.* Small sample effects in time series analysis: A new asymptotic theory and a new estimate // *Ann. Statist.* 1988. V. 16. N. 2. С. 808–841.