Журнал Белорусского государственного университета. Математика. Информатика. 2019;2:34–39
Journal of the Belarusian State University. Mathematics and Informatics. 2019;2:34–39

УДК 519.24

ПОСТРОЕНИЕ ОЦЕНОК СПЕКТРАЛЬНЫХ ПЛОТНОСТЕЙ С ЗАДАННОЙ ТОЧНОСТЬЮ ПО ПЕРЕСЕКАЮЩИМСЯ ИНТЕРВАЛАМ НАБЛЮДЕНИЙ

Н. В. СЕМЕНЧУК[1)]

[1)]*Гродненский государственный университет им. Янки Купалы, ул. Ожешко, 22, 230023, г. Гродно, Беларусь*

Представлен новый метод определения числа интервалов разбиений и количества наблюдений в них при построении оценок спектральных плотностей стационарных случайных процессов по пересекающимся интервалам наблюдений с заданной точностью на основе асимптотических результатов, полученных для скорости сходимости первого момента в предположении, что спектральная плотность удовлетворяет условию Липшица. Изучены два случая: с единичным и произвольным окном просмотра данных. В результате предложен алгоритм построения оценок по пересекающимся интервалам наблюдений с заданной точностью. Данный алгоритм апробирован на модельных примерах для случайных процессов AR(4) посредством использования окна просмотра данных Рисса, Бохнера, Парзена. Предложенный способ будет полезен при анализе данных в виде стационарных случайных процессов с помощью непараметрических методов спектрального анализа в автоматизированном режиме.

Ключевые слова: спектральная плотность; стационарный случайный процесс; смещение оценки; оценки по пересекающимся интервалам наблюдений с заданной точностью.

CONSTRUCTION OF ESTIMATES OF SPECTRAL DENSITIES WITH A GIVEN ACCURACY OVER INTERSECTING INTERVALS OF OBSERVATIONS

N. V. SEMENCHUK[a]

[a]*Yanka Kupala State University of Grodno, 22 Ažeška Street, Hrodna 230023, Belarus*

The article proposes a new method for determining the number of splitting intervals and the number of observations in them when building estimates of the spectral densities of stationary random processes with a given accuracy over intersecting observation intervals based on asymptotic results, obtained for the first moment of convergence rate under the assumption that the spectral density satisfies the Lipschitz condition. Two cases are considered: with a single and arbitrary data taper. As a result, an algorithm is proposed for constructing estimates for intersecting intervals of observations with a given accuracy. This algorithm was tested on model examples for random AR(4) processes, using data taper of Riesz, Bochner, Parzen. The proposed method will be useful to the researcher in analyzing data in the form of stationary random processes using non-parametric methods of spectral analysis in an automated mode.

Keywords: spectral density; stationary random process; estimate bias; estimates for overlapping observation intervals with a given accuracy.

Образец цитирования:
Семенчук НВ. Построение оценок спектральных плотностей с заданной точностью по пересекающимся интервалам наблюдений. *Журнал Белорусского государственного университета. Математика. Информатика.* 2019;2:34–39.
https://doi.org/10.33581/2520-6508-2019-2-34-39

For citation:
Semenchuk NV. Construction of estimates of spectral densities with a given accuracy over intersecting intervals of observations. *Journal of the Belarusian State University. Mathematics and Informatics*. 2019;2:34–39. Russian.
https://doi.org/10.33581/2520-6508-2019-2-34-39

Автор:
Наталья Владимировна Семенчук – доцент кафедры фундаментальной и прикладной математики факультета математики и информатики.

Author:
Natalia V. Semenchuk, associate professor at the department of fundamental and applied mathematics, faculty of mathematics and informatics.
senata155@gmail.com

Введение

Статья посвящена решению задачи выбора числа разбиений при построении оценок спектральных плотностей по пересекающимся интервалам наблюдений с заданным смещением. Эта задача возникла из необходимости структурно анализировать данные в виде стационарных случайных процессов различной длины в автоматизированном режиме, в частности для спектрального анализа электрокардиограмм, динамики производственных показателей [1] и др. Отметим, что графики оценок спектральной плотности помогают выявить почти периодические компоненты спектра и их месторасположение, а также могут быть полезны при выборе модели. При построении оценок спектральных плотностей обычно применяются периодограммные методы [2], в основе которых лежит квадрат модуля преобразования Фурье конечной реализации исследуемого процесса. Для получения состоятельных оценок спектральных плотностей, как правило, используется метод сглаживания периодограмм спектральными окнами. Обработка случайного процесса с помощью функций окна просмотра данных и спектрального окна [3], как и при построении оценок спектральных плотностей по пересекающимся и непересекающимся интервалам наблюдений, применяется для улучшения статистических свойств оценок его спектральной плотности. При этом основное назначение окна просмотра данных – уменьшить величину смещения, а спектрального окна или процедуры усреднения периодограмм – дисперсию в спектральных оценках [4].

Материалы и методы исследования

Пусть число наблюдений T за стационарным случайным процессом $X(t),\ t \in Z$, представимо в виде [1]

$$T = S(N - M) + M,$$

где S – число пересекающихся интервалов разбиения длиной N; M принимает целочисленные значения, $S \le M < N$ (S не зависит от T).

На l-м интервале построим расширенную периодограмму

$$I^{(h)}_{l(N-M)}(\lambda) = \frac{1}{2\pi H_2^{(T)}(0)} d_T(\lambda) d_T(-\lambda), \tag{1}$$

где

$$d_T(\lambda) = \sum_{t=0}^{T-1} h_T(t) X(t) e^{-i\lambda t};\ \ H_k^{(T)}(\lambda) = \sum_{t=0}^{T-1} \left(h_T(t)\right)^k e^{-i\lambda t}, \tag{2}$$

$h_T(t) = h\left(\frac{t}{T}\right)$, $h : [0, 1] \to \mathbf{R}$ – функция окна просмотра данных; $l, k, T \in \mathbf{N}$.

В качестве оценки спектральной плотности $f(\lambda),\ \lambda \in \Pi = [-\pi, \pi]$, рассмотрим статистику вида

$$\hat{f}_T(\lambda) = \frac{1}{S} \sum_{l=1}^{S} I^{(h)}_{l(N-M)}(\lambda), \tag{3}$$

построенную путем усреднения расширенных периодограмм по S пересекающимся интервалам наблюдений.

Исследуем скорость сходимости смещения оценки спектральной плотности, построенную по пересекающимся интервалам наблюдений.

Лемма 1 [3]. *Для любого действительного* $\alpha \in (0, 1]$, $T = 1, 2, \ldots$ *справедливо неравенство*

$$\int_{\Pi} |x|^{\alpha} \Phi_T(x)\, dx \le K(\alpha, T),$$

где $\Phi_T(x)$ *– ядро Фейера, задаваемое равенством*

$$\Phi_T(x) = \frac{1}{2\pi T} \frac{\sin \frac{Tx}{2}}{\sin \frac{x}{2}},\ x \in \Pi; \tag{4}$$

$$K(\alpha, T) = \begin{cases} \dfrac{2^{\alpha}\pi}{T^{\alpha}\left(1-\alpha^2\right)} - \dfrac{\pi}{T(1-\alpha)}, \textit{ если } 0 < \alpha < 1, \\ \dfrac{2\pi\left(\dfrac{1}{2} + \ln\dfrac{\pi T}{2}\right)}{T}, \textit{ если } \alpha = 1. \end{cases} \tag{5}$$

Лемма 2 [4]. *Пусть окно просмотра данных – функция с вариацией ограниченной постоянной $V > 0$. Тогда для любого действительного $\alpha \in (0, 1]$, $T = 1, 2, \ldots$ справедливо неравенство*

$$\int\limits_{\Pi} |x|^{\alpha} \Phi_2^{(T)}(x)dx \le C_1(\lambda)^2 V^2 R(\alpha, T),$$

где $\Phi_2^{(T)}(x)$ – ядро, задаваемое равенством

$$\Phi_2^{(T)}(x) = \frac{\left|H_1^{(T)}(x)\right|\left|H_1^{(T)}(-x)\right|}{2\pi H_2^{(T)}(0)}; \tag{6}$$

$$R(\alpha, T) = \begin{cases} \dfrac{4T^{1-\alpha}}{2\pi H_2^{(T)}(0)\left(1-\alpha^2\right)} - \dfrac{2\pi^{\alpha-1}}{2\pi H_2^{(T)}(0)(1-\alpha)}, \textit{ если } 0 < \alpha < 1, \\ \dfrac{2\ln(\pi T) + 1}{2\pi H_2^{(T)}(0)}, \textit{ если } \alpha = 1; \end{cases} \tag{7}$$

$$C_1(\lambda) = \begin{cases} 1, \textit{ если } |\lambda| \le \dfrac{1}{T}, \\ \pi, \textit{ если } \dfrac{1}{T} < |\lambda| \le \pi, \end{cases} \tag{8}$$

$H_1^{(T)}(x)$, $H_2^{(T)}(0)$ определены формулой (2).

Учитывая выражения для первого момента и доказательство свойства асимптотической несмещенности оценки (3), полученные для единичного окна просмотра данных в [1], в настоящей работе исследуем скорость сходимости смещения оценки (3) для единичного и произвольного окна просмотра данных.

Рассмотрим спектральные плотности $f(\lambda) \in Lip_{\alpha}(L)$, где $Lip_{\alpha}(L)$ – множество функций, удовлетворяющих условию Липшица порядка $\alpha \in (0, 1]$, $0 < L < \infty$.

В случае единичного окна просмотра данных для скорости сходимости смещения оценки (3) имеем следующий результат.

Теорема 1. *Для любого действительного $\alpha \in (0, 1]$, $f(\lambda) \in Lip_{\alpha}(L)$ и $T = 1, 2, \ldots$ справедливо неравенство*

$$\left|M\hat{f}_T(\lambda) - f(\lambda)\right| \le LK(\alpha, N),$$

где $K(\alpha, N)$ определяется соотношением (5).

Доказательство. Из [3] известно, что математическое ожидание статистики, заданной соотношением (3), равно

$$M\hat{f}^{(T)}(\lambda) = \int\limits_{\Pi} f(x+\lambda)\Phi_N(x)dx,$$

где $\Phi_N(x)$ определяется формулой (4); $\lambda \in \Pi$. Далее, используя свойства ядра Фейера $\Phi_N(x)$ [1], для смещения оценки (3) имеем

$$\left|M\hat{f}^{(T)}(\lambda) - f(\lambda)\right| = \left|\int\limits_{\Pi} f(x+\lambda)\Phi_N(x)dx - f(\lambda)\right| \le \int\limits_{\Pi} \left|f(x+\lambda) - f(\lambda)\right|\Phi_N(x)dx.$$

Так как $f(\lambda) \in Lip_{\alpha}(L)$, $\lambda \in \Pi$, с учетом леммы 1 окончательно получим

$$\left| M\hat{f}^{(T)}(\lambda) - f(\lambda) \right| \le \int_{\Pi} |x|^{\alpha} \Phi_N(x)dx \le LK(\alpha, N).$$

Теорема доказана.

Теорема 2. *Если окно просмотра данных – функция с вариацией ограниченной постоянной $V > 0$, $f(\lambda) \in Lip_{\alpha}(L)$, то для любого действительного $\alpha \in (0, 1]$, $T = 1, 2, \ldots$ справедливо неравенство*

$$\left| M\hat{f}_N(\lambda) - f(\lambda) \right| \le LC_1(\lambda)^2 V^2 R(\alpha, N),$$

где $H_2^{(N)}(0)$, $R(\alpha, N)$, $C_1(\lambda)$ задаются соотношениями (2), (7), (8) соответственно.

Доказательство. Из [3] известно, что математическое ожидание статистики, заданной соотношением (3), имеет вид

$$M\hat{f}^{(T)}(\lambda) = \int_{\Pi} f(x + \lambda)\Phi_2^{(N)}(x)dx,$$

где $\Phi_2^{(N)}(x)$ – ядро, определенное формулой (6); $\lambda \in \Pi$.

Далее, используя свойства ядра $\Phi_2^{(N)}(x)$, с учетом леммы 2 имеем

$$\left| M\hat{f}^{(T)}(\lambda) - f(\lambda) \right| = \left| \int_{\Pi} f(x + \lambda)\Phi_2^{(N)}(x)dx - f(\lambda) \right| \le \int_{\Pi} \left| f(x + \lambda) - f(\lambda) \right| \Phi_2^{(N)}(x)dx.$$

Так как $f(\lambda) \in Lip_{\alpha}(L)$, $\lambda \in \Pi$, окончательно получим

$$\left| M\hat{f}^{(T)}(\lambda) - f(\lambda) \right| \le L\int_{\Pi} |x|^{\alpha} \Phi_2^{(N)}(x)dx \le LC_1(\lambda)^2 V^2 R(\alpha, N).$$

Теорема доказана.

Отметим, что для оценки спектральной плотности (3) асимптотическая дисперсия в S раз меньше, чем асимптотическая дисперсия расширенной периодограммы, построенной по T наблюдениям [3].

Замечание. Численные исследования, проведенные в работах [4; 5], показывают, что для различных окон просмотра данных справедлива оценка

$$\int_{\Pi} |x|^{\alpha} \Phi_2^{(T)}(x)dx \le R(\alpha, T),$$

более точная, чем доказанная в лемме 2, что дает возможность строить оценки спектральных плотностей с заданной точностью даже при небольшой длине реализации временного ряда.

Далее в статье будем рассматривать случай $\alpha = 1$ как наиболее часто встречающийся при решении практических задач. Отметим также, что, как правило, вариация для большинства упомянутых окон просмотра данных равна 1; константа Липшица $L = \max\left| f'(x) \right|$ при $\alpha = 1$. Решение задачи оценки производных спектральных плотностей по конечной реализации наблюдений за случайным процессом можно найти в работах В. Г. Алексеева и И. Г. Журбенко [6–8].

Для наглядности приведем график функции (7) с окном просмотра данных Рисса, Бохнера, Парзена (рис. 1). Ряд исследований в [5] показал, что данное окно позволяет строить расширенные периодограммы с наименьшим смещением.

Далее для построения оценок рекомендуется использовать следующий алгоритм (при фиксированном $\varepsilon > 0$ и числе наблюдений T).

Шаг 1. Выбрать окно просмотра данных (согласно исследованиям, приведенным в [5], рекомендуется использовать окно просмотра данных Рисса, Бохнера, Парзена).

Шаг 2. Подобрать число наблюдений N для каждого интервала разбиения так, чтобы $R(1, N) \le \varepsilon$.

Шаг 3. Выбрать S так, чтобы дисперсия конечной оценки была меньше, чем дисперсия периодограммы, построенной по всему количеству наблюдений T.

Шаг 4. Определить число наблюдений в пересечении интервалов разбиения:

$$M = \frac{S \cdot N - T}{S - 1}.$$

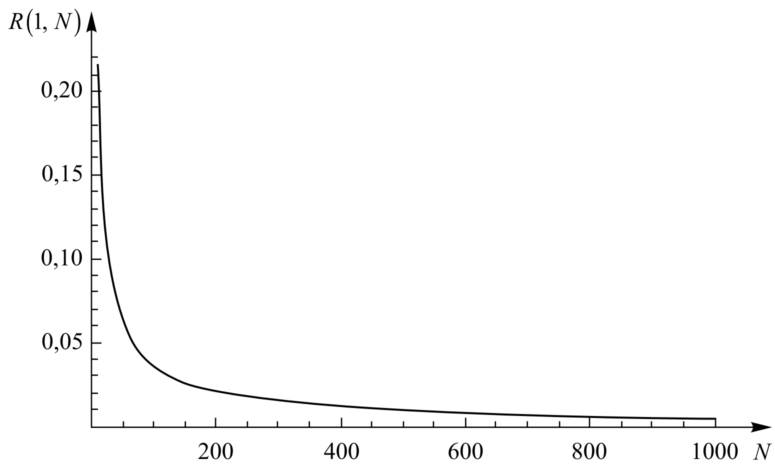

Рис. 1. График функции $R(1, N)$ с окном просмотра данных Рисса, Бохнера, Парзена

Fig. 1. Function graph with a data taper of Riesz, Bochner, Parzen

Шаг 5. На каждом разбиении построить расширенную периодограмму (1).

Шаг 6. Построить оценку спектральной плотности усреднением периодограмм по количеству интервалов по формуле (3).

Результаты и их обсуждение

Экспериментально проиллюстрируем результаты, полученные в теореме 1. Сгенерируем 20 реализаций процесса авторегрессии 4-го порядка с заданными пиками при $a_1 = 0{,}217\,37$; $a_2 = 0{,}817\,14$; $a_3 = 0{,}176\,07$; $a_4 = -0{,}656\,1$ и при $T = 128$ и $T = 1024$. Далее вычислим оценки спектральной плотности (3) при $N = 256$ и $M = 128$ с единичным окном просмотра данных, а также при $N = 32$ и $M = 16$ с окном просмотра данных Рисса, Бохнера, Парзена, используя предложенный алгоритм.

Найдем приближение математического ожидания оценки спектральной плотности методом Монте-Карло, а также вычислим значение верхней границы в неравенствах, полученных в теоремах 1 и 2 для величины смещения. На рис. 2 и 3 представлены результаты теорем 1 и 2 при $\alpha = 1$ с учетом изменения значения константы Липшица для теоретической спектральной плотности смоделированного процесса. На рис. 4 и 5 приведены построенные оценки и теоретическая спектральная плотность.

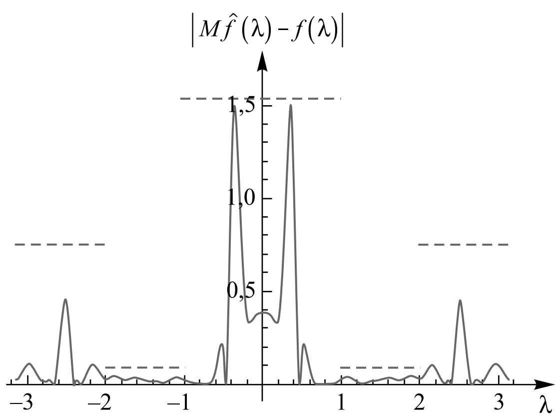

Рис. 2. Иллюстрация результата теоремы 1 при $T = 1024$, $N = 256$

Fig. 2. Illustration of the result of theorem 1 at $T = 1024$, $N = 256$

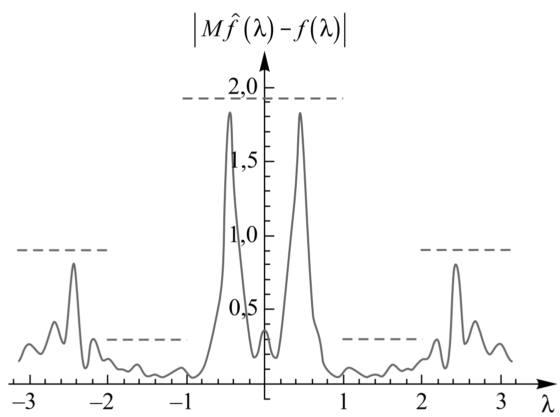

Рис. 3. Иллюстрация результата теоремы 2 при $T = 128$, $N = 32$ с окном просмотра данных Рисса, Бохнера, Парзена

Fig. 3. Illustration of the result of theorem 2 at $T = 128$, $N = 32$ with Riesz, Bochner, Parzen data taper

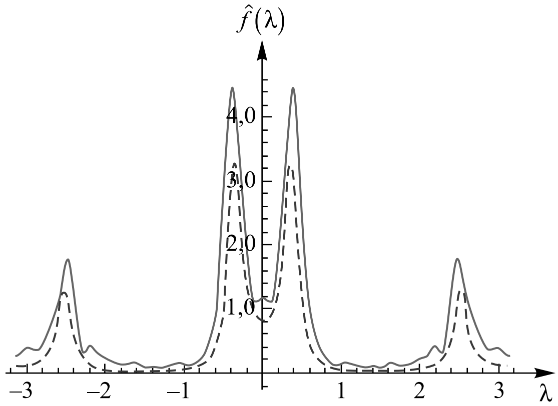

Рис. 4. Теоретическая спектральная плотность и ее оценка (3) с единичным окном просмотра данных ($T = 1024$, $N = 256$)

Fig. 4. Theoretical spectral density and its estimate (3) with a single data taper ($T = 1024$, $N = 256$)

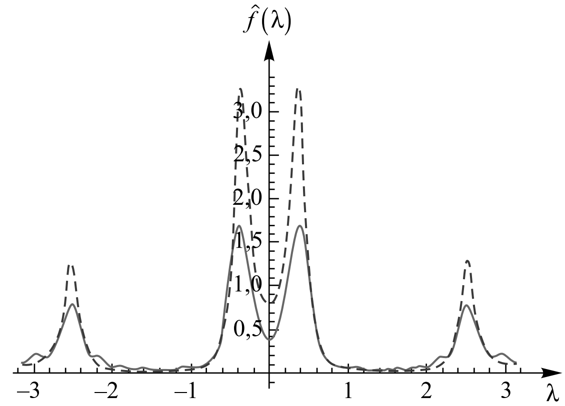

Рис. 5. Теоретическая спектральная плотность и ее оценка (3) с окном просмотра данных Рисса, Бохнера, Парзена ($T = 128$, $N = 16$)

Fig. 5. Theoretical spectral density and its estimate (3) with Riesz, Bochner, Parzen data taper ($T = 128$, $N = 16$)

Результаты, приведенные в настоящей статье, позволяют выбирать параметры оценивания (число интервалов разбиения и количество наблюдений в интервале) при построении оценок спектральных плотностей по пересекающимся интервалам наблюдений, основываясь на скорости сходимости моментов этих оценок.

Библиографические ссылки

1. Сурмач АИ, Семенчук НВ. Методы анализа данных при помощи состоятельных оценок спектральных плотностей. В: *Теория вероятностей, случайные процессы, математическая статистика и приложения. Сборник научных статей международной конференции, посвященной 80-летию профессора, доктора физико-математических наук Г. А. Медведева; 23–26 февраля 2015 г.; Минск, Беларусь.* Минск: РИВШ; 2015. с. 305–310.

2. Бриллинджер Д. *Временные ряды. Обработка данных и теория.* Москва: Мир; 1980. 536 с.

3. Труш НН. *Асимптотические методы статистического анализа временных рядов.* Минск: БГУ; 1999. 218 с.

4. Семенчук НВ, Труш НН. Сравнительный анализ некоторых периодограммных оценок спектральных плотностей. *Труды Института математики.* 2007;15(2):90–103.

5. Семенчук НВ. О выборе оптимального окна просмотра данных. *Веснік Гродзенскага дзяржаўнага ўніверсітэта імя Янкі Купалы. Серыя 2. Матэматыка. Фізіка. Інфарматыка, вылічальная тэхніка і кіраванне.* 2012;2(129):61–66.

6. Журбенко ИГ. *Спектральный анализ временных рядов.* Москва: МГУ; 1982. 168 с.

7. Алексеев ВГ. О вычислении спектров стационарных случайных процессов по выборкам большого объема. *Проблемы передачи информации.* 1980;16(1):42–49.

8. Алексеев ВГ. Об оценках спектральных плотностей некоторых моделей стационарных случайных процессов. *Проблемы передачи информации.* 1985;21(2):42–49.

References

1. Surmach AI, Semenchuk NV. [Methods of data analysis using consistent spectral density estimates]. In: *Teoriya veroyatnostei, sluchainye protsessy, matematicheskaya statistika i prilozheniya. Sbornik nauchnykh statei mezhdunarodnoi konferentsii, posvyashchennoi 80-letiyu professora, doktora fiziko-matematicheskikh nauk G. A. Medvedeva; 23–26 fevralya 2015 g.; Minsk, Belarus'* [Probability theory, random processes, mathematical statistics and applications. Proceedings of the International conference in honor of 80 years jubilee of professor, doctor of physical and mathematical sciences Gennady Medvedev; 2015 February 23–26; Minsk, Belarus]. Minsk: Respublikanskii institut vysshei shkoly; 2015. p. 305–310. Russian.

2. Brillindzher D. *Vremennye ryady. Obrabotka dannykh i teoriya* [Time series. Data processing and theory]. Moscow: Mir; 1980. 536 p. Russian.

3. Troush NN. *Asimptoticheskie metody statisticheskogo analiza vremennykh ryadov* [Asymptotic methods of statistical analysis of time series]. Minsk: Belarusian State University; 1999. 218 p. Russian.

4. Semenchuk NV, Troush NN. The comparative analysis of some periodogram's estimates of spectral density. *Trudy Instituta matematiki.* 2007;15(2):90–103. Russian.

5. Semenchuk NV. [About choosing the optimal data taper]. *Vesnik Grodzenskaga dzjarzhawnaga wniversiteta imja Janki Kupaly. Seryja 2. Matjematyka. Fizika. Infarmatyka, vylichal'naja tjehnika i kiravanne.* 2012;2(129):61–66. Russian.

6. Zhurbenko IG. *Spektral'nyi analiz vremennykh ryadov* [Spectral analysis of time series]. Moscow: Moscow State University; 1982. 168 p. Russian.

7. Alekseev VG. [On the calculation of the spectra of stationary random processes for samples of large volume]. *Problemy peredachi informatsii*. 1980;16(1):42–49. Russian.

8. Alekseev VG. [Estimates of the spectral densities of some models of stationary random processes]. *Problemy peredachi informatsii.* 1985;21(2):42–49. Russian.

Статья поступила в редколлегию 22.02.2019.
Received by editorial board 22.02.2019.