

АСИМПТОТИЧЕСКИЕ СВОЙСТВА КЕПСТРА ГАУССОВСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

В. А. Волошко

НИИ прикладных проблем математики и информатики

Минск, Беларусь

E-mail: valeravoloshko@yandex.ru

Получены асимптотические разложения центральных моментов оценок кепстральных коэффициентов гауссовского стационарного временного ряда с условием регулярности при возрастающей длительности наблюдения. Сделаны качественные выводы о вероятностных свойствах оценок кепстральных коэффициентов.

Ключевые слова: кепстр, лог-периодограмма, стационарность, гауссовость.

ВВЕДЕНИЕ

Кепстральный анализ применяется в задачах распознавания речи, сейсмологии, обработки сигналов, робастного прогнозирования временных рядов [8, 9]. Он основан на преобразовании Фурье логарифма спектра мощности наблюдаемого временного ряда.

В литературе доказаны основные свойства отсчетов периодограммы [3], такие как асимптотическая некоррелированность для стационарных в широком смысле временных рядов и асимптотическая независимость и χ^2_2 -распределенность для временных рядов, стационарных в узком смысле [3, 4] (при растущей длительности наблюдения $T \rightarrow \infty$). Также установлены некоторые свойства оценок кепстральных ко-

эффицентов при определенных условиях регулярности, например их асимптотическая некоррелированность и равнодисперсность [5].

В данной работе исследуются моменты статистических оценок кепстральных коэффициентов гауссовского временного ряда при выполнении условия регулярности на спектральную плотность.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПОСТАНОВКА ЗАДАЧИ

Пусть $\{x_t\}, t \in \mathbb{Z}$, – гауссовский стационарный временной ряд с нулевым математическим ожиданием $E\{x_t\} = 0$, спектральной плотностью (сокращенно, спектром) $S(\lambda)$ и лог-спектром $L(\lambda) = \ln S(\lambda)$, определенный на вероятностном пространстве (Ω, F, P) :

$$S(\lambda) = S(-\lambda) = \sum_{\tau \in \mathbb{Z}} \sigma_\tau \cos(\tau\lambda), \quad L(\lambda) = L(-\lambda) = \sum_{\tau \in \mathbb{Z}} l_\tau \cos(\tau\lambda), \quad \lambda \in \Pi = [-\pi, \pi],$$

$$\sigma_\tau = \sigma_{-\tau} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\lambda) \cos(\tau\lambda) d\lambda, \quad l_\tau = l_{-\tau} = \frac{1}{2\pi} \int_{-\pi}^{\pi} L(\lambda) \cos(\tau\lambda) d\lambda, \quad (1)$$

где коэффициенты ряда Фурье для $S(\lambda)$ – ковариации $\{\sigma_\tau\}$, а для $L(\lambda)$ – так называемые (в теории обработки сигналов) кепстральные коэффициенты [4, 5]. Введем условие регулярности, характеризующее асимптотику l_τ при $\tau \rightarrow \infty$:

$$\chi = \overline{\lim_{\tau \rightarrow +\infty}} \sqrt[3]{|l_\tau|} < 1. \quad (2)$$

При выполнении условия (2) ряды Лорана $\sum_{\tau \in \mathbb{Z}} \sigma_\tau z^\tau$ и $\sum_{\tau \in \mathbb{Z}} l_\tau z^\tau$ сходятся в кольце $\chi < |z| < 1/\chi$, а $L(\lambda)$ – бесконечно дифференцируемая и ограниченная функция. В частности, условию (2) удовлетворяют все временные ряды авторегрессии и скользящего среднего конечного порядка без нулей и полюсов в спектре [2].

Построим дискретное преобразование Фурье (ДПФ) временного ряда $x_1, \dots, x_T$ длины $T \in \mathbb{N}$:

$$y_\lambda = u_\lambda + iv_\lambda = \frac{1}{\sqrt{T}} \sum_{t=1}^T x_t e^{i(t-t_0)\lambda}, \quad \beta_\lambda = \ln |y_\lambda|^2, \quad \lambda \in \mathbb{R}, \quad (3)$$

где i – мнимая единица, $t_0 = (T+1)/2$ – среднее значение индексов времени наблюдения $1, 2, \dots, T$, которое введено в (3) для упрощения дальнейших математических выкладок. Согласно [2], статистика $|y_\lambda|^2$, определяемая (3), называется периодограммой, а β_λ – лог-периодограммой [10]. Построим подстановочные («plug-in») статистические оценки для кепстральных коэффициентов $\{l_\tau\}$:

$$\hat{l}_\tau = \frac{1}{T} \sum_{\lambda=1}^T \beta_{\tilde{\lambda}} \cos(\tau\tilde{\lambda}), \quad \tilde{\lambda} = \frac{2\pi\lambda}{T}, \quad \tau \in \Lambda, \quad (4)$$

где $\Lambda = \left\{0, 1, \dots, \left\lfloor \frac{T}{2} \right\rfloor\right\}$, $\lfloor z \rfloor, \lceil z \rceil$ – ближайшие к $z \in \mathbb{R}$ целые снизу и сверху соответственно.

МОМЕНТЫ ДПФ $\{y_\lambda\}$

Нам понадобится ряд вспомогательных результатов для ДПФ $\{y_\lambda\}$. Введем комплекснозначные функции:

$$Q(\lambda) = \sum_{\tau \in \mathbb{Z}} \sigma_\tau e^{i\tau\lambda} = S(\lambda) + iH(\lambda), \quad Q_T(\lambda) = \sum_{\tau=1-T}^{T-1} \sigma_\tau e^{i\tau\lambda} = S_T(\lambda) + iH_T(\lambda). \quad (5)$$

Теорема 1. Если имеет место математическая модель (1)–(3), то для ковариаций $e_{v,\mu} = E\{y_v y_\mu\}$, $A_{v,\mu} = E\{u_v u_\mu\}$, $B_{v,\mu} = E\{v_v v_\mu\}$ справедливы следующие выражения:

$$e_{v,\mu} = \frac{\operatorname{Im}(e^{i\lambda T} \bar{q}_T)}{T \sin \lambda} = \frac{\operatorname{Im}(e^{i\lambda T} \bar{q})}{T \sin \lambda} + \Delta_{v,\mu}, \quad A_{v,\mu}, B_{v,\mu} = \frac{e_{-v,\mu} \pm e_{v,\mu}}{2}, \quad E\{u_v v_\mu\} = 0,$$

где вектор $(q, q_T, \lambda) = \frac{1}{2} \sum_{x=v,\mu} (Q(x), Q_T(x), x)$, $|\Delta_{v,\mu}| \leq \sum_{\tau > T} |\sigma_\tau| \left| \left(\frac{\tau}{T} - 1 \right) \right|$; черта означает комплексное сопряжение.

Из теоремы 1 следует, что процессы $\{u_\lambda\}$ и $\{v_\lambda\}$ некоррелированы, а в силу гауссовости (вытекающей из гауссовости x_t и линейности преобразования (3)) и независимы. Введем обозначения: $1\{A\}$ – индикатор события A ; $\Lambda_0 = \{0, T/2\} \cap \mathbb{Z}$;

$$a_{v,\mu} = \frac{A_{v,\mu}}{\sqrt{A_{v,v} A_{\mu,\mu}}}, \quad b_{v,\mu} = \frac{B_{v,\mu}}{\sqrt{B_{v,v} B_{\mu,\mu}}}$$

– корреляционные функции для процессов $\{u_\lambda\}$ и $\{v_\lambda\}$,

$$B_{\lambda,\lambda} = \rho_\lambda (1 + \omega_\lambda), \quad A_{\lambda,\lambda} = \rho_\lambda (1 - \omega_\lambda), \quad \sqrt{B_{\lambda,\lambda}} = \vartheta_\lambda (1 + \theta_\lambda), \quad \sqrt{A_{\lambda,\lambda}} = \vartheta_\lambda (1 - \theta_\lambda).$$

На основе (5) построим функции:

$$\Psi(v, \mu) = \frac{H(v) + H(\mu)}{2 \sin \frac{v+\mu}{2}}, \quad \Psi_T(v, \mu) = \frac{H_T(v) + H_T(\mu)}{2 \sin \frac{v+\mu}{2}}.$$

Следствие 1. В условиях теоремы 1 для $v, \mu \in \mathbb{Z}$

$$e_{\tilde{v}, \tilde{\mu}} = \varepsilon_{v+\mu} S_T(\tilde{\mu}) - \frac{\Psi_T(\tilde{v}, \tilde{\mu})}{T} (-1)^{v+\mu} = \varepsilon_{v+\mu} S(\tilde{\mu}) - \frac{\Psi(\tilde{v}, \tilde{\mu})}{T} (-1)^{v+\mu} + \Delta_{\tilde{v}, \tilde{\mu}},$$

где $\tilde{\lambda}$ определяется (4) и $\varepsilon_n = 1\{\frac{n}{T} \in \mathbb{Z}\} (-1)^{\frac{n(T+1)}{T}}$. Кроме того, если выполнено условие регулярности (2), то равномерно по $v, \mu \in \Lambda, v \neq \mu$, справедливы следующие асимптотические разложения при $T \rightarrow \infty$ ($\chi < \chi_+ < 1$):

$$e_{-\tilde{v}, \tilde{\mu}} = \left(S - \frac{H'}{T} \right) (\tilde{\mu}) + O(\chi_+^T), \quad \omega_{\tilde{\mu}} = \begin{cases} T^{-1} (H/S \sin)(\tilde{\mu}) + O(T^{-2}), & \mu \notin \Lambda_0, \\ (-1)^{1+2\mu/T}, & \mu \in \Lambda_0, \end{cases}$$

$$a_{\tilde{v}, \tilde{\mu}}, b_{\tilde{v}, \tilde{\mu}} = (-1)^{v+\mu+1} \frac{\Psi(\tilde{v}, -\tilde{\mu}) \pm \Psi(\tilde{v}, \tilde{\mu})}{T \sqrt{2^\varepsilon S(\tilde{v}) S(\tilde{\mu})}} + O\left(\frac{1}{T^2}\right), \quad \varepsilon = \varepsilon_{2v}^2 + \varepsilon_{2\mu}^2.$$

МОМЕНТЫ ЛОГ-ПЕРИОДОГРАММЫ $\{\beta_\lambda\}$

Теорема 2. Пусть $k > 1$, $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, степени $n_1, \dots, n_k \in \{0, 1, \dots\}$, $k_1 = \sum_{j=1}^k \delta_{1, n_j}$ и ε – норма вектора, составленного из $a_{\lambda_r, \lambda_q}, b_{\lambda_r, \lambda_q}$ для $r, q = 1, \dots, k$, $r < q$. Тогда справедливы следующие асимптотические соотношения при $\varepsilon \rightarrow 0$ ($\lambda_{1,2} = \nu, \mu$):

$$E \left\{ \prod_{j=1}^k \dot{\beta}_{\lambda_j}^{n_j} \right\} = \prod_{j=1}^k E \left\{ \dot{\beta}_{\lambda_j}^{n_j} \right\} + O(\varepsilon^{\max\{2, k_1\}}),$$

$$E \{ \dot{\beta}_\nu \dot{\beta}_\mu \} = \frac{a_{\nu, \mu}^2 (1 - \theta_\nu)(1 - \theta_\mu) + b_{\nu, \mu}^2 (1 + \theta_\nu)(1 + \theta_\mu)}{2} + O(\varepsilon^3). \quad (6)$$

В случае, когда $\theta_\nu = \theta_\mu = -1$, момент $E \{ \dot{\beta}_\nu \dot{\beta}_\mu \} = 2 \arcsin^2 a_{\nu, \mu}$. Формула (6) тогда примет вид: $2 \arcsin^2 a_{\nu, \mu} = 2a_{\nu, \mu}^2 + O(\varepsilon^3)$, $\varepsilon = |a_{\nu, \mu}| \rightarrow 0$.

Следствие 3. Если выполнено условие регулярности (2), то для каждого $k > 1$ и набора степеней $n_1, \dots, n_k \in \{0, 1, \dots\}$, $k_1 = \sum_{j=1}^k 1 \{n_j = 1\}$, следующие асимптотические соотношения выполняются при $T \rightarrow +\infty$ равномерно по всем наборам попарно различных $\lambda_1, \dots, \lambda_k \in \Lambda$ ($\lambda_1 = \forall \lambda_2 = \mu$):

$$\varepsilon = O\left(\frac{1}{T}\right), \quad E \left\{ \prod_{j=1}^k \dot{\beta}_{\lambda_j}^{n_j} \right\} = \prod_{j=1}^k E \left\{ \dot{\beta}_{\lambda_j}^{n_j} \right\} + O\left(\frac{1}{T^{\max\{2, k_1\}}}\right),$$

$$E \{ \dot{\beta}_{\tilde{\nu}} \dot{\beta}_{\tilde{\mu}} \} = \frac{\Psi^2(\tilde{\nu}, -\tilde{\mu}) + \Psi^2(\tilde{\nu}, \tilde{\mu})}{T^2 S(\tilde{\nu}) S(\tilde{\mu})} + O\left(\frac{1}{T^3}\right),$$

где ε из теоремы 2 вычисляются для значений $\tilde{\lambda}_1, \dots, \tilde{\lambda}_k$.

МОМЕНТЫ ОЦЕНОК КЕПСТРАЛЬНЫХ КОЭФФИЦИЕНТОВ $\{\hat{l}_\tau\}$

Обозначим: $1 \{m | n\}$ – индикатор делимости n на m ; $\bar{1} \{m | n\} = 1 - 1 \{m | n\}$; $g_n = 1 + 1 \{2 | T\} (-1)^n$;

$$c_\tau = \int_{\Pi^n} \frac{c(\lambda)}{(2\pi)^n} \prod_{j=1}^n \cos(\tau_j \lambda_j) d\lambda, \quad c: \Pi^n \rightarrow \mathbb{R}, \quad \tau \in \mathbb{Z}^n, -$$

многомерный косинус-коэффициент функции $c(\lambda)$; N – разбиение множества индексов $\{1, \dots, n\}$ на k непустых подмножеств $N_1, \dots, N_k$ мощности $n_1, \dots, n_k$ соответственно ($n_1 + \dots + n_k = n$), $k_1 = \sum_{j=1}^k 1 \{n_j = 1\}$.

Теорема 3. При условии регулярности (2) для моментов статистических оценок кепстральных коэффициентов (4) равномерно по $\tau, \tau', \tau_j \in \Lambda$ справедливы следующие асимптотические разложения при $T \rightarrow +\infty$:

$$E\{\hat{l}_\tau\} = -\gamma\delta_{0,\tau} + l_\tau - \frac{1}{T} \left(\frac{H'}{S} \right)_\tau - \frac{1}{2T^2} \left(\frac{(H')^2 + \frac{H^2}{2\sin^2}}{S^2} \right)_\tau - \frac{g_\tau}{T} \ln 2 + O\left(\frac{1}{T^3}\right), \quad (7)$$

$$E\left\{\prod_{j=1}^n \hat{l}_{\tau_j}\right\} = \frac{1}{T^{\lfloor n/2 \rfloor}} \left(O\left(\frac{1}{T}\right) + F_n \sum_{N:k=\lfloor n/2 \rfloor}^{k_1=0} \prod_{j=1}^{\lfloor n/2 \rfloor} R(\tau_{r \in N_j}) \right), \quad n > 1, \quad (8)$$

$$E\{\hat{l}_\tau \hat{l}_{\tau'}\} = \frac{\pi^2}{6T} \left(R(\tau, \tau') + \frac{g_{\tau+\tau'}}{T} \right) + \frac{2}{T^2} \left(\frac{\Psi^2(v, \mu)}{S(v)S(\mu)} \right)_{\tau, \tau'} + O\left(\frac{1}{T^3}\right),$$

где $R(\tau_1, \dots, \tau_r) = \sum_{\tau=\tau_1 \pm \dots \pm \tau_r} 1\{T \mid \tau\}$, $F_n = (-2A)^p (\pi^2/6)^q$, $p = \bar{1}\{2 \mid n\}$, $q = (n-3p)/2$.

Учитывая постоянную часть смещения в (7) для $\tau = 0$, мы можем использовать статистику $\hat{l}_0 + \gamma$ для оценивания коэффициента l_0 , который называется энтропией [1]. С такой поправкой оценки (4) при условии регулярности (2) в асимптотике растущей длительности наблюдения $T \rightarrow \infty$ из (7) асимптотически не смещены и, как следует из (8), сходятся в среднеквадратичном к кепстральным коэффициентам, что влечет их состоятельность. Также отметим следующее свойство робастности: главный член смешанного момента центрированных оценок (8) не зависит от спектральной плотности временного ряда, а только от набора индексов. Кроме того, для четных порядков n этот главный член согласно (8) вычисляется так же, как и совместный момент n гауссовских в совокупности случайных величин: по всем разбиениям n индексов на пары суммируются произведения $n/2$ моментов второго порядка. Поэтому главные члены моментов четного порядка в (8) в точности равны соответствующим моментам независимых гауссовских случайных величин $\xi_{i \in \Lambda}$ с дисперсиями $D\{\xi_i\} = 2^{1\{i \in \Lambda_0\}} \pi^2/6T$. Асимметрия распределения оценок (4) проявляется в ненулевых главных членах моментов нечетного порядка, где возникает постоянная Аперри.

ЛИТЕРАТУРА

1. Amari, S. Methods of Information Geometry / Amari S. NY, Oxford University Press, 2000. 206 p.
2. Voloshko, V. A. Statistical Forecasting Based on Bloomfield Exponential Model/ V. A. Voloshko, Yu. S. Kharin // Computer Data Analysis and Modelling 2010. Proceedings of the 9th International Conference CDAM 2010. Minsk, 2010. Vol. 1. P. 268–271.
3. Андерсон, Т. В. Статистический анализ временных рядов / Т. В. Андерсон. М.: Мир, 1976. 756 с.
4. Бриллинджер, Д. Временные ряды. Обработка данных и теория / Д. Бриллинджер. М.: Мир, 1980. 536 с.
5. Merhav, N. On the Asymptotic Statistical Behavior of Empirical Cepstral Coefficients / N. Merhav // IEEE Transactions On Signal Processing. 1993. Vol. 41. № 5. P. 1990–1993.
6. Bloomfield, P. An Exponential Model for the Spectrum of a Scalar Time Series / P/Bloomfield // Biometrika. 1973. Vol. 60. P. 217–226.
7. Бейтмен, Г. Высшие трансцендентные функции. / Г. Бейтмен, А. Эрдейи. М.: Наука, 1965. Т. 1. 294 с.
8. Харин, Ю. С. Оптимальность и робастность в статистическом прогнозировании / Ю. С. Харин. Минск: БГУ, 2008. 263 с.
9. Kharin, Yu. S. Robust estimation of AR coefficients under simultaneously influencing outliers and missing values / Yu. S. Kharin, V. A. Voloshko // J. Statist. Plann. Inference. 2011. Vol. 141. I. 9. P. 3276–3288.

10. *Arteche, J.* Bootstrap-based bandwidth choice for log-periodogram regression / J. Arteche., J. Orbe. // J. Time Series Analysis. 2009. Vol. 6. P. 591–617.
 11. *Шабат, Б. В.* Введение в комплексный анализ / Б. В. Шабат. М.: Наука, 1987. 577 с.
-

|

(

i