АСИМПТОТИЧЕСКИЕ СВОЙСТВА КЕПСТРА ГАУССОВСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

В. А. Волошко

НИИ Прикладных проблем математики и информатики БГУ
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 $T \to \infty$). Также установлены некоторые свойства оценок кепстральных коэффициентов 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):

$$S(\lambda)=S(-\lambda)=\sum_{\tau\in\mathbb{Z}}\sigma_\tau\cos(\tau\lambda),\quad L(\lambda)=L(-\lambda)=\sum_{\tau\in\mathbb{Z}}l_\tau\cos(\tau\lambda),\quad \lambda\in\Pi=[-\pi,\pi],$$
$$\sigma_\tau=\sigma_{-\tau}=\frac{1}{2\pi}\int_{-\pi}^{\pi}S(\lambda)\cos(\tau\lambda)d\lambda,\quad l_\tau=l_{-\tau}=\frac{1}{2\pi}\int_{-\pi}^{\pi}L(\lambda)\cos(\tau\lambda)d\lambda,\tag{1}$$

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 $\tau\to\infty$:

$$\chi=\overline{\lim_{\tau\to+\infty}}\sqrt[\tau]{|l_\tau|}<1.\tag{2}$$

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 $T\in\mathbb{N}$:

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

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 $\{l_\tau\}$:

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

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:

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

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

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

где вектор $\left(q,q_T,\lambda\right)=\frac{1}{2}\sum_{x=\nu,\mu}\left(Q(x),Q_T(x),x\right)$, $\left|\Delta_{\nu,\mu}\right|\le\sum_{\tau>T}\left|\sigma_\tau\right|\left(\frac{\tau}{T}-1\right)$*; 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; $\Lambda_0=\{0,T/2\}\bigcap\mathbb{Z}$;

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

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

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

На основе (5) построим функции:

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

Следствие 1. *В условиях Теоремы 1 для* $\nu,\mu\in\mathbb{Z}$

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

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

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

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

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

Теорема 3. *Пусть* $k>1$, $\lambda_1,\ldots,\lambda_k\in\mathbb{R}$, *степени* $n_1,\ldots,n_k\in\{0,1,\ldots\}$, $k_1=\sum_{j=1}^k\delta_{1,n_j}$ *и* ε *– норма вектора, составленного из* $a_{\lambda_r,\lambda_q},b_{\lambda_r,\lambda_q}$ *для* $r,q=1,\ldots,k$, $r<q$. *Тогда справедливы следующие асимптотические соотношения при* $\varepsilon\to0$ *(*$\lambda_{1,2}=\nu,\mu$*):*

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

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

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

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

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

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

где ε *из Теоремы 3 вычисляются для значений* $\tilde{\lambda}_1,\ldots,\tilde{\lambda}_k$.

МОМЕНТЫ ОЦЕНОК КЕПСТРАЛЬНЫХ КОЭФФИЦИЕНТОВ $\left\{\widehat{l}_\tau\right\}$

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

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

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

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

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

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

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

где $R(\tau_1,\ldots,\tau_r)=\sum\limits_{\tau=\tau_1\pm\ldots\pm\tau_r}1\{T\mid\tau\}$, $F_n=(-2A)^p\left(\pi^2/6\right)^q$, $p=\overline{1}\{2\mid n\}$, $q=(n-3p)/2$.

Учитывая постоянную часть смещения в (7) для $\tau=0$, мы можем использовать статистику $\widehat{l}_0+\gamma$ для оценивания коэффициента l_0, который называется энтропией [1]. С такой поправкой оценки (4) при условии регулярности (2) в асимптотике растущей длительности наблюдения $T\to\infty$ 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 Апери.

ЛИТЕРАТУРА

1. Amari S. Methods of Information Geometry / Amari S. NY, Oxford University Press, 2000. 206 p.
2. Voloshko V.A., Kharin Yu.S. Statistical Forecasting Based on Bloobfield Exponential Model // Computer Data Analysis and Modelling 2010. Proceedings of the 9th International Conference CDAM 2010. Minsk, 2010. V. 1. P. 268-271.
3. Андерсон Т.В. Статистический анализ временных рядов / Андерсон Т.В. М: Мир, 1976. 756 с.
4. Бриллинджер Д. Временные ряды. Обработка данных и теория / Бриллинджер Д. М: Мир, 1980. 536 с.
5. Merhav N. On the Asymptotic Statistical Behavior of Empirical Cepstral Coefficients / Merhav N // IEEE Transactions On Signal Processing. 1993. V. 41. N. 5. P. 1990-1993.
6. Bloomfield P. An Exponential Model for the Spectrum of a Scalar Time Series / Bloomfield P // Biometrika. 1973. V. 60. P. 217-226.
7. Бейтмен Г., Эрдейи А. Высшие трансцендентные функции – том 1 / Бейтмен Г., Эрдейи А. М: Наука, 1965. 294 с.
8. Харин Ю.С. Оптимальность и робастность в статистическом прогнозировании / Харин Ю.С. Минск: БГУ, 2008. 263 с.
9. Kharin Yu.S., Voloshko V.A. Robust estimation of AR coefficients under simultaneously influencing outliers and missing values / Kharin Yu.S., Voloshko V.A // J. Statist. Plann. Inference. 2011. V. 141. I. 9. P. 3276-3288.
10. Arteche J., Orbe J. Bootstrap-based bandwidth choice for log-periodogram regression / Arteche J., Orbe J // J. Time Series Analysis. 2009. V. 6. P. 591-617.
11. Шабат Б.В. Введение в комплексный анализ / Шабат Б.В. М: Наука, 1987. 577 с.