

ОБ ОЦЕНКЕ ВРЕМЕННОГО ЗАПАЗДЫВАНИЯ

к.ф.-м.н. В.Н. Шаляпин, Л.А. Егорова, Ю.В. Левадный
(представил д.ф.-м.н. В.К. Иванов)

Основным методом определения временного запаздывания между проявлениями одного и того же случайного процесса, пришедшего к наблюдателю по различным траекториям, является поиск максимума взаимной корреляционной функции. В работе рассчитано распределение вероятности оценки взаимной корреляционной функции в зависимости от количества учитываемых пар точек, временного смещения между реализациями и степени их корреляции.

Введение. Во многих задачах распространения волн исходный сигнал $s(t)$ может попадать к наблюдателю по нескольким траекториям. Поскольку длины траекторий различны, временные изменения сигнала, повторяющиеся в каждом из каналов, усиливаясь на различные величины, регистрируются в приемнике с некоторым запаздыванием τ относительно друг друга. Кроме исходного сигнала в каждом из трактов присутствуют свои собственные добавочные сигналы $n(t)$, вызванные условиями прохождения сигнала вдоль данной трассы и ошибками измерений. Для двухканального распространения в дискретном представлении, когда моменту t соответствует i -е измерение, а сдвигу на τ отвечает смещение на j отсчетов, соотношение между входным s и выходными x и y сигналами принимает вид [1]:

$$x(i) = s(i) + n_x; \quad y(i) = \alpha \cdot s(i - j) + n_y. \quad (1)$$

где через α обозначено относительное усиление сигнала.

В большинстве приложений можно считать, что шумы p_x и p_y не коррелируют между собой и с исходным процессом. Кроме того, без потери общности, можно принять, что они имеют нулевое среднее значение.

Можно привести множество примеров распространения сигналов, описываемых формулой (1), из акустики, геофизики, радиофизики или астрономии [2]. Основным методом определения временной задержки является построение выборочной взаимной ковариационной функции двух наблюдаемых рядов [3]:

$$R_{xy}^{\epsilon}(k, N) = \frac{1}{N} \sum_{i=1}^N x(i)y(i+k). \quad (2)$$

Смещение между двумя сериями k , при которых $R_{xy}^{\epsilon}(k, N)$ максимально, принимается в качестве оценки временного запаздывания.

Качественное поведение случайной функции $R_{xy}^{\epsilon}(k, N)$ несложно понять, если заметить, что при больших N она стремится к взаимной ковариационной функции $R_{xy}(k)$, которая в свою очередь выражается через автоковариационную функцию исходного процесса:

$$\lim_{N \rightarrow \infty} R_{xy}^{\epsilon}(k, N) = R_{xy}(k) = \langle x(i)y(i+k) \rangle = \langle [s(i) + n_x] \cdot [\alpha \cdot s(i+k-j) + n_y] \rangle = \quad (3)$$

$$= \alpha \langle s(i)s(i+k-j) \rangle = \alpha R_{xx}(k-j).$$

В последнем равенстве использовано предположение о стационарности процесса. При k , равном временному запаздыванию j , автоковариационная функция $R_{xx}(0)$ достигает своего максимального значения, равного $\langle s^2 \rangle + \langle s \rangle^2$. Отклонение k от j приводит к уменьшению корреляции между сериями. При достаточно больших смещениях $|k-j|$ значения $x(i)$ и $y(i+k)$ перестают коррелировать между собой и становятся независимыми, а взаимная ковариационная функция спадает до своего минимального значения $R_{xy}(\infty) = \langle s \rangle^2$.

Если возможно проведение многих измерений с равномерными временными интервалами, то оценка ковариационной функции не представляет больших сложностей [4]. Напротив, при небольшом количестве измеренных пар точек N поведение случайной функции $R_{xy}^{\epsilon}(k, N)$ заметно отличается от усредненного значения $R_{xy}(k)$ и может носить довольно изрезанный, неравномерный характер. Случайные выбросы при произвольных значениях k вполне могут превысить значение максимума при $k = j$. Для анализа подобных ситуаций было предложено несколько различных методов, основанных на дискретной корреляции [5], линейной фильтрации [6, 7], спектральном анализе [8] и др. Однако полное определение вероятности попадания на истинный максимум возможно только на основе анализа распределения вероятностей случайных величин $R_{xy}^{\epsilon}(k, N)$ для каждой фиксированной пары значений (k, N) . Эти же распределения вероятностей используются и для разделения случайных величин при соседних k , т.е. для определения точности оценки временного запаздывания.

1. Распределение вероятности в максимуме взаимной корреляционной функции. Прежде всего, сделаем некоторые самые общие до-

пущения о характере исходного процесса $s(i)$. Предположим, что он является стационарным нормальным процессом с нулевым средним и единичной дисперсией. Отметим, что при нулевом среднем и единичной дисперсии ковариационная функция совпадает с нормированной корреляционной функцией. В отсутствии шума при попадании на истинную временную задержку каждое слагаемое в сумме (2) представляет собой нормальную величину с плотностью распределения вероятности

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (4)$$

умноженную саму на себя. Несложно найти плотность распределения вероятности $h(z)$ такой величины: $z(x) = x^2$. Учитывая, что обратная функция $x(z) = \pm\sqrt{z}$ двузначна, имеем (см., например [9])

$$h(z) = \frac{1}{\sqrt{2\pi}} z^{-1/2} e^{-z/2}, \quad 0 \leq z < \infty. \quad (5)$$

В сумму (2) входит N слагаемых с одинаковой плотностью вероятности $h(z)$. Для того, чтобы найти вероятность распределения суммы можно произвести N -кратную свертку слагаемых. Однако, удобнее воспользоваться характеристической функцией $\phi(u)$ для распределения $h(z)$:

$$\phi(u) = \int_{-\infty}^{+\infty} e^{iuz} h(z) dz = \frac{1}{2\pi} \int_0^{+\infty} e^{iuz} z^{-1/2} e^{-z/2} dz = \frac{1}{\sqrt{1-2iu}}. \quad (6)$$

Характеристическая функция суммы N одинаковых слагаемых представляет собой просто произведение N характеристических функций каждого слагаемого, т.е.

$$\phi_N(u) = [\phi(u)]^N = (1-2iu)^{-N/2}. \quad (7)$$

Обратное фурье-преобразование от $\phi_N(u)$ дает плотность распределения суммы из N квадратов нормально распределенных величин

$$\begin{aligned} h_N(z) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_N(u) e^{-iuz} du = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-iuz}}{(1-2iu)^{N/2}} du = \\ &= \frac{1}{2^{N/2} \Gamma(N/2)} z^{\frac{N}{2}-1} e^{-z/2}, \quad 0 \leq z < \infty. \end{aligned} \quad (8)$$

Функция (8) является хорошо известным распределением χ_N^2 с N степенями свободы. В данном случае оно описывает плотность распределения вероятности произведения случайного процесса самого на себя или, другими словами, при нулевом сдвиге, т.е. в точке максимума. Первый и второй моменты M_1 , M_2 распределения (8) удобно найти непосредствен-

ным дифференцированием характеристической функции (7):

$$M_1 = \frac{1}{i} \frac{\partial \phi_N}{\partial u} \Big|_{u=0} = N; \quad (9)$$

$$M_2 = \left(\frac{1}{i} \right)^2 \frac{\partial^2 \phi_N}{\partial u^2} \Big|_{u=0} = N(N+2). \quad (10)$$

Отсюда дисперсия распределения равна $\sigma^2 = M_2 - M_1^2 = 2N$.

2. Распределение вероятности в минимуме взаимной корреляционной функции. При больших временных сдвигах относительно друг друга случайные величины $x(i)$ и $y(i+k)$ становятся независимыми. Каждое слагаемое в сумме (2) представляет собой произведение двух нормальных случайных величин $Z = X \cdot Y$. Функция распределения $H(z)$ случайной величины Z может быть выражена через совместную плотность вероятности $f(x, y)$:

$$H(z) = \int\limits_{D(z)} f(x, y) dx dy. \quad (11)$$

Интегрирование проводится по области $D(z)$, в которой $z \leq x \cdot y$. Учитывая функциональную зависимость $z(x, y)$, имеем

$$H(z) = \int_{-\infty}^0 dx \int_{z/x}^{+\infty} f(x, y) dy + \int_0^{+\infty} dx \int_{-\infty}^{z/x} f(x, y) dy = 2 \int_0^{+\infty} dx \int_{-\infty}^{z/x} f(x, y) dy. \quad (12)$$

Плотность вероятности $h(z)$, являющаяся производной от функции распределения $H(z)$, получается с помощью дифференцирования интеграла (12) по верхнему пределу

$$h(z) = \frac{dH(z)}{dz} = 2 \int_0^{+\infty} \frac{dx}{x} \frac{1}{2\pi} e^{-x^2/2 - z^2/2x^2} \quad (13)$$

или, выполнив интегрирование,

$$h(z) = \frac{1}{\pi} K_0(z), \quad (14)$$

где через $K_\nu(z)$ обозначена модифицированная функция Бесселя третьего рода или функция Макдональда порядка ν . Выражение (14) описывает плотность распределения вероятности произведения двух независимых нормальных величин, с характеристической функцией

$$\phi(u) = 2 \int_0^{+\infty} h(z) \cos uz dz = \frac{1}{\sqrt{1+u^2}}. \quad (15)$$

Характеристическая функция суммы из N одинаковых слагаемых получается возведением в N -ую степень выражения (15):

$$\phi_N(u) = (1 + u^2)^{-N/2}. \quad (16)$$

Итоговая плотность распределения суммы вычисляется обратным фурье-преобразованием от (16):

$$h_N(z) = \frac{1}{\pi} \int_0^{+\infty} \phi(u) \cos uz \, dz = \frac{1}{\sqrt{\pi} \Gamma(N/2)} \left(\frac{|z|}{2} \right)^{\frac{N-1}{2}} K_{(n-1)/2}(z). \quad (17)$$

Из выражения (16) для характеристической функции легко получаем

среднее $M_1 = \frac{1}{i} \frac{\partial \phi_N}{\partial u} \Big|_{u=0} = 0$ и дисперсию $\sigma^2 = M_2 = \left(\frac{1}{i} \right)^2 \frac{\partial^2 \phi_N}{\partial u^2} \Big|_{u=0} = N$ распределения (17).

3. Распределение вероятности в произвольной точке взаимной корреляционной функции. В предыдущих двух разделах были рассмотрены два предельных распределения вероятностей в точках максимума и минимума корреляционной функции. Перейдем теперь к рассмотрению распределения в произвольной точке.

Допустим, что две серии наблюдений сдвинуты относительно друг друга на некоторую величину. Каждое значение сдвига характеризуется своим коэффициентом корреляции ρ , который является мерой статистической зависимости между двумя рядами при данном смещении. При сдвиге равном величине запаздывания, коэффициент корреляции $\rho \rightarrow 1$ и по отсчету в ряде X можно однозначно с вероятностью 1 предсказать соответствующее значение в ряде Y (в нашем упрощенном случае эти значения просто равны друг другу). При очень большом сдвиге измеренные значения становятся статистически независимыми и $\rho \rightarrow 0$. Конечный, но не очень большой сдвиг, характеризуется своим значением ρ , лежащем в диапазоне $-1 \leq \rho \leq 1$. Конкретный вид зависимости ρ от смещения определяется автокорреляционной функцией исходного процесса $s(t)$ и величиной шумовых добавок $n(x)$ и $n(y)$ в (1).

Совместная плотность вероятности одновременного обнаружения зависимых значений x и y в двух сериях, определяется теперь двумерным нормальным распределением вероятности

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}[x^2 - 2xy + y^2]\right). \quad (18)$$

Распределение произведения $z = x \cdot y$ можно получить по формуле (13):

$$h(z) = 2 \cdot \frac{1}{2\pi} \sqrt{1-\rho^2} \int_0^{+\infty} \exp\left(-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho z + z^2/x^2]\right) dx. \quad (19)$$

После выполнения интегрирования находим плотность распределения вероятности произведения двух нормальных коррелированных величин

$$h(z) = \frac{e^{\frac{\rho z}{1-\rho^2}}}{\pi\sqrt{1-\rho^2}} K_0\left(\frac{z}{1-\rho^2}\right). \quad (20)$$

Распределения (5) и (14) являются частными случаями распределения (20) при значениях коэффициента корреляции $\rho = 0$ и 1 соответственно. Формулу (5) можно получить из (20), воспользовавшись асимптотическим представлением функции Макдоналда K_ν при больших значениях аргумента

$$K_\nu \rightarrow \sqrt{\frac{\pi}{2x}} e^{-x}, \quad x \gg 1, \nu. \quad (21)$$

Перейдем теперь к вычислению характеристической функции $\phi(u)$ от $h(z)$:

$$\begin{aligned} \phi(u) &= \int_{-\infty}^{+\infty} e^{iuz} h(z) dz = \int_0^{+\infty} [e^{-iuz} h(-z) + e^{iuz} h(z)] dz = \\ &= \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^{+\infty} \left[e^{-iuz - \frac{\rho z}{1-\rho^2}} K_0\left(\frac{z}{1-\rho^2}\right) + e^{iuz + \frac{\rho z}{1-\rho^2}} K_0\left(\frac{z}{1-\rho^2}\right) \right] dz = \\ &= \frac{2}{\pi\sqrt{1-\rho^2}} \int_0^{+\infty} \cos\left[\left(u - \frac{i\rho}{1-\rho^2}\right)z\right] K_0\left(\frac{z}{1-\rho^2}\right) dz. \end{aligned} \quad (22)$$

Окончательно получаем

$$\phi(u) = \frac{1}{\sqrt{(1-\rho^2)u^2 - 2i\rho u + 1}}. \quad (23)$$

Снова отметим соответствующие предельные случаи (6) и (15) для $\rho \rightarrow 1; 0$. Характеристическая функция суммы равна

$$\phi_N(u) = [(1-\rho^2)u^2 - 2i\rho u + 1]^{-N/2}. \quad (24)$$

Обратное фурье-преобразование от него дает искомую функцию:

$$h_N(z) = \frac{1}{\sqrt{\pi} \Gamma(N/2) \sqrt{1-\rho^2}} \left(\frac{|z|}{2}\right)^{\frac{N-1}{2}} e^{\frac{\rho z}{1-\rho^2}} K_{(N-1)/2}\left(\frac{z}{1-\rho^2}\right); \quad (25)$$

$$M_1 = \rho N; \quad M_2 = (1+\rho^2)N + \rho^2 N^2; \quad \sigma^2 = (1+\rho^2)N. \quad (26)$$

Итоговое выражение (25) является основным результатом данной работы. Оно полностью описывает распределение вероятностей суммы (2)

для любого количества учитываемых пар измерений N и произвольных сдвигах серий относительно друг друга. Величина сдвига выражается через коэффициент корреляции ρ . Конкретный вид зависимости сдвига от ρ может быть установлен из априорных соображений или определяться экспериментально по измерению автокорреляционной функции процесса.

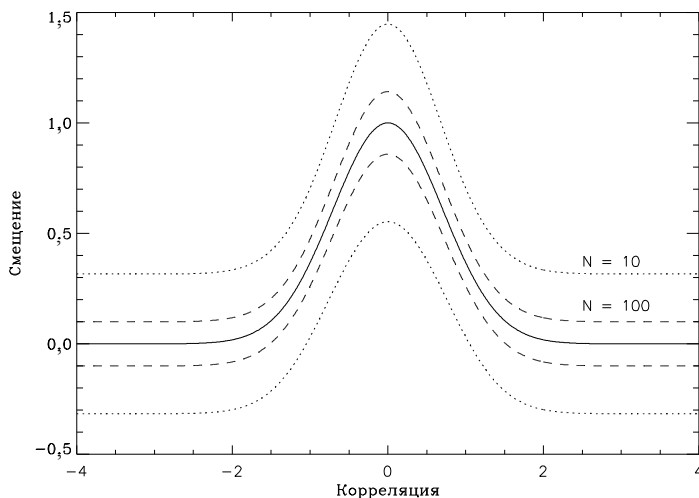


Рис. 1. Зависимость степени корреляции от смещения между сериями (пунктирной линией показана дисперсия при учете 10 пар точек, а штриховой — при 100 измерениях)

Обсуждение. На рис. 1 показана зависимость средней величины и дисперсии функции распределения (25) от смещения между сериями измерений Δt при различном количестве учитываемых пар измерений N . Для сравнения результатов с разными N все величины нормированы на максимальное значение корреляции, достигаемое при нулевом смещении, т.е. на N . Для определенности закон изменения корреляции от смещения выбран как $\rho = \exp(-\Delta t)^2$. Из графика видно, что нормированные средние не зависят от количества учитываемых пар и совпадают с абсолютно точным значением при $N \rightarrow \infty$. В то же время относительная дисперсия с увеличением N уменьшается, позволяя все точнее оценивать корреляционную зависимость. Примечательно, что дисперсия в максимуме корреляционной функции в $\sqrt{2}$ раз выше, чем в минимуме.

Аналитическое выражение, получаемое по формуле (25), может быть использовано при исследовании целого ряда вопросов, связанных с поиском сдвига между наблюдаемыми рядами. Например, при оценке вероят-

ности возникновения ложных максимумов корреляционной функции или для расчета точности определения положения истинного максимума. Однако необходимо все время учитывать условия, при которых оно было получено. Основным предположением, которое закладывалось при выводе (25), было условие *независимости* между суммированными парами слагаемых. Такое условие выполняется только в ограниченном круге задач, например, если величина запаздывания намного меньше или значительно больше расстояния между парами измерений. В тех же случаях, когда эти величины сравнимы по значению, возможно возникновение существенных корреляций между измеренными парами и результаты данной работы можно применять только с определенной степенью осторожности для получения качественных оценок поведения корреляционной функции.

ЛИТЕРАТУРА

1. Carter G.S., Knapp C.H., Nuttal A.H. Estimation of the magnitude-squared coherence via overlapped fast Fourier transform processing // *IEEE transactions on Audio and Electroacoustics*. – 1973. – AU-21. – P. 337.
2. Refsdal S. On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect // *Monthly Notices Roy. Astron. Soc.* – 1964. – V. 128. – P. 307.
3. Яглом А.М. Корреляционная теория стационарных функций. – Л.: Гидрометеоиздат, 1981. – 237 с.
4. Бендат Дж. Основы теории случайных шумов и ее применение. – М.: Наука, 1965. – 304 с.
5. Edelson R.A., Krolik J.H. The discrete correlation function - A new method for analyzing unevenly sampled variability data // *Astroph. J.* – 1988. – V. 333. – P. 646.
6. Press, W.H., Rybicki, G.B., Hewitt, J.N. The time delay of gravitational lens 0957+561 // *Astroph. J.* – 1992. – V. 385. – P. 404.
7. Rybicki G.B., Press W.H. Interpolation, realization, and reconstruction of noisy, irregularly sampled data // *Astroph. J.* – 1992. – V. 398. – P.169.
8. Scargle J.D. Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data // *Astroph. J.* – 1989. – V. 343. – P. 874.
9. Худсон Д. Статистика для физиков. – М.: Мир, 1967. – 232 с.

Поступила 11.01.2005

ШАЛЯПИН Вячеслав Николаевич, канд. физ.-мат. наук, научный сотрудник ИРЭ НАНУ. Окончил Харьковский национальный университет. Область научных исследований – радиолокация, обработка информации. E-mail: vshal@ire.kharkov.ua

ЕГОРОВА Любовь Алексеевна, мл. научн. сотр. ИРЭ НАНУ. Окончила Национальный аэрокосмический университет «ХАИ». Область научных исследований – дистанционное зондирование Земли с аэрокосмических носителей.

ЛЕВАДНЫЙ Юрий Владимирович, аспирант ИРЭ НАНУ. Окончил ХНУ радиотехники в 2004 году. Область научных исследований – радиотехника СВЧ.