

АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ОЦЕНКИ ВАРИОГРАММЫ ГАУССОВСКОГО СЛУЧАЙНОГО ПРОЦЕССА

Т. В. Цеховая

Белорусский государственный университет

Минск, Беларусь

E-mail: Tsekhavaya@bsu.by

Построена оценка вариограммы внутренне стационарного гауссовского случайного процесса с дискретным временем. Найдены выражения для первых двух моментов и семиинвариантов высших порядков исследуемой статистики. При условии, что ряд из семивариограмм абсолютно сходится, исследовано асимптотическое поведение ковариации, дисперсии и семиинвариантов высших порядков оценки вариограммы. Найдено предельное распределение изучаемой статистики.

Ключевые слова: случайный процесс, гауссовский случайный процесс, стационарный случайный процесс, внутренняя стационарность, вариограмма, оценка.

В настоящее время проблема построения и изучения статистических свойств оценок основных характеристик временных рядов является достаточно актуальной [1-7]. Интерес представляет статистический анализ оценок ковариационной функции и вариограммы, поскольку они являются функциями, определяющими зависимость между наблюдениями временного ряда.

При решении многих прикладных задач исследователи пытаются свести изучение исходных наблюдений к теории нормальных случайных процессов. Это связано с тем, что вышеуказанная теория разработана наиболее полно. Заметим также, что гауссовские случайные процессы принадлежат классу внутренне стационарных случайных процессов, которые адекватно описывают многие математические модели в геологии, экологии, эпидемиологии и т.д.

Случайный процесс $X(s)$, $s \in Z = \{0, \pm 1, \pm 2, \dots\}$, называется внутренне стационарным, если справедливы следующие равенства:

$$M\{X(s_1) - X(s_2)\} = 0, \quad D\{X(s_1) - X(s_2)\} = 2\gamma(s_1 - s_2),$$

где $2\gamma(s_1 - s_2)$ – вариограмма рассматриваемого процесса, $s_1, s_2 \in Z$. Заметим, что функция $\gamma(s)$, $s \in Z$, – называется семивариограммой.

В данной работе обобщаются вопросы исследования статистических свойств оценок вариограммы нормальных случайных процессов, рассмотренные в [3, 5, 7].

Пусть $X(s)$, $s \in Z$, – внутренне стационарный гауссовский случайный процесс с нулевым математическим ожиданием, дисперсией σ^2 и неизвестной вариограммой

$$2\gamma(h) = D\{X(s+h) - X(s)\}, \quad s, h \in Z.$$

Заметим, что

$$\{X(s+h) - X(s)\}^2 = 2\gamma(h) \cdot \chi_1^2,$$

где χ_1^2 – случайная величина, распределенная по закону хи-квадрат с одной степенью свободы. Легко показать, что

$$\begin{aligned} M\{X(s+h) - X(s)\}^2 &= 2\gamma(h), \\ D\{X(s+h) - X(s)\}^2 &= 2\{2\gamma(h)\}^2. \end{aligned}$$

Предположим далее, что $X(1), \dots, X(n)$ – n последовательных, полученных через равные промежутки времени наблюдений за процессом $X(s)$, $s \in Z$. В качестве оценки вариограммы рассмотрим статистику вида

$$2\tilde{\gamma}(h) = \frac{1}{n-h} \sum_{s=1}^{n-h} (X(s+h) - X(s))^2, \quad (1)$$

$h = \overline{0, n-1}$. Положим $\tilde{\gamma}(-h) = \tilde{\gamma}(h)$, $h = \overline{0, n-1}$, и $\tilde{\gamma}(h) = 0$ для $|h| \geq n$.

Найдем выражения для первых двух моментов оценки вариограммы (1).

Теорема 1. Для оценки $2\tilde{\gamma}(h)$, заданной равенством (1), имеют место следующие соотношения:

$$\begin{aligned} M\{2\tilde{\gamma}(h)\} &= 2\gamma(h), \\ \text{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} &= \\ &= \frac{2}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} \{\gamma(t+h_1-s) + \gamma(t-s-h_2) - \gamma(t+h_1-s-h_2) - \gamma(t-s)\}^2, \end{aligned} \quad (2)$$

$$D\{2\tilde{\gamma}(h)\} = \frac{2}{(n-h)^2} \sum_{t,s=1}^{n-h} \{\gamma(t-s+h) + \gamma(t-s-h) - 2\gamma(t-s)\}^2, \quad (3)$$

где $\gamma(h)$, $h \in Z$, – семивариограмма процесса $X(s)$, $s \in Z$, $h, h_1, h_2 = \overline{0, n-1}$.

Доказательство. Из определения вариограммы и свойств математического ожидания, первое утверждение теоремы вытекает очевидным образом:

$$M\{2\tilde{\gamma}(h)\} = \frac{1}{n-h} \sum_{s=1}^{n-h} M(X(s+h) - X(s))^2 = \frac{1}{n-h} \sum_{s=1}^{n-h} 2\gamma(h) = 2\gamma(h).$$

Используя определение ковариации, подставляя вместо $2\tilde{\gamma}(h)$ ее выражение в явном виде, запишем

$$\text{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} = \frac{1}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} \text{cov}\{(X(t+h_1) - X(t))^2, (X(s+h_2) - X(s))^2\}.$$

Учитывая определение коэффициента корреляции, утверждения лемм 1 [4, с. 78] и 3 [4, с. 79], получим

$$\begin{aligned} \text{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} &= \\ &= \frac{1}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} \text{corr}\{(X(t+h_1) - X(t))^2, (X(s+h_2) - X(s))^2\} \times \\ &\quad \times \sqrt{D(X(t+h_1) - X(t))^2 D(X(s+h_2) - X(s))^2} = \\ &= \frac{2\{2\gamma(h_1)\}\{2\gamma(h_2)\}}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} \left\{ \frac{\gamma(t+h_1-s) + \gamma(t-s-h_2) - \gamma(t+h_1-s-h_2) - \gamma(t-s)}{\sqrt{2\gamma(h_1)}\sqrt{2\gamma(h_2)}} \right\}^2, \end{aligned}$$

откуда следует требуемое равенство (2) для ковариации.

Соотношение (3) для дисперсии оценки вариограммы $2\tilde{\gamma}(h)$ нетрудно получить из (2), если положить $h_1 = h_2 = h$. Таким образом,

$$\begin{aligned} D\{2\tilde{\gamma}(h)\} &= \frac{2\{2\gamma(h)\}^2}{(n-h)^2} \sum_{t,s=1}^{n-h} \left\{ \frac{\gamma(t-s+h) + \gamma(t-s-h) - 2\gamma(t-s)}{2\gamma(h)} \right\}^2 = \\ &= \frac{2}{(n-h)^2} \sum_{t,s=1}^{n-h} \{\gamma(t-s+h) + \gamma(t-s-h) - 2\gamma(t-s)\}^2. \end{aligned}$$

Теорема доказана.

Нетрудно видеть, что статистика (1) является несмещенной оценкой для вариограммы $2\gamma(h)$.

Исследуем асимптотическое поведение моментов второго порядка построенной оценки $2\tilde{\gamma}(h)$, $h, = \overline{0, n-1}$.

Теорема 2. Если имеет место соотношение

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < \infty, \quad (4)$$

то

$$\begin{aligned} \lim_{n \rightarrow \infty} (n - \min\{h_1, h_2\}) \text{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} &= \\ = 2 \sum_{m=-\infty}^{+\infty} \{\gamma(m-h_2) + \gamma(m+h_1) - \gamma(m+h_1-h_2) - \gamma(m)\}^2, \end{aligned} \quad (5)$$

$$\lim_{n \rightarrow \infty} (n-h) D\{2\tilde{\gamma}(h)\} = 2[(2\gamma(h))^2 + 2 \sum_{m=1}^{+\infty} \{\gamma(m-h) + \gamma(m+h) - 2\gamma(m)\}^2], \quad (6)$$

где $\gamma(h)$, $h \in Z$, – семивариограмма процесса $X(s)$, $s \in Z$, $h, h_1, h_2 = \overline{0, n-1}$.

Доказательство. Рассмотрим равенство (2). Пусть $h_1 > h_2$. Сделаем замену переменных суммирования: $t = t$, $t - s = m$, тогда

$$\begin{aligned} \text{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} &= \frac{2}{n-h_2} \left[\sum_{m=-(n-h_2-1)}^{n-h_1-1} \{\gamma(m+h_1) + \gamma(m-h_2) - \gamma(m+h_1-h_2) - \gamma(m)\}^2 - \right. \\ &\quad \left. - \frac{2}{n-h_1} \sum_{m=1}^{n-h_1-1} m \{\gamma(m+h_1) + \gamma(m-h_2) - \gamma(m+h_1-h_2) - \gamma(m)\}^2 \right]. \end{aligned}$$

Аналогично рассуждаем для случая $h_1 < h_2$.

$$\begin{aligned} \text{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} &= \frac{2}{n-h_1} \left[\sum_{m=-(n-h_2-1)}^{n-h_1-1} \{\gamma(m+h_1) + \gamma(m-h_2) - \gamma(m+h_1-h_2) - \gamma(m)\}^2 + \right. \\ &\quad \left. + \frac{2}{n-h_2} \sum_{m=-(n-h_2-1)}^{-1} m \{\gamma(m+h_1) + \gamma(m-h_2) - \gamma(m+h_1-h_2) - \gamma(m)\}^2 \right]. \end{aligned}$$

Объединяя полученные результаты, учитывая (4), вытекает требуемое предельное соотношение (5) для ковариации.

Равенство (6) для дисперсии оценки вариограммы $2\tilde{\gamma}(h)$ нетрудно получить из (5), если положить $h_1 = h_2 = h$.

Следствие. Из теоремы 2 вытекает, что $\lim_{n \rightarrow \infty} D\{2\tilde{\gamma}(h)\} = 0$, $h = \overline{0, n-1}$.

В силу первого утверждения теоремы 1 и вышеуказанного следствия получаем, что $2\tilde{\gamma}(h)$ является состоятельной в среднеквадратическом смысле оценкой для вариограммы $2\gamma(h)$, $h \in Z$.

Исследуем асимптотическое поведение семиинвариантов высших порядков статистики (1).

Теорема 3. Пусть справедливо (4), тогда для оценки $2\tilde{\gamma}(h)$, $h = \overline{0, n-1}$, задаваемой равенством (1), имеет место

$$\lim_{n \rightarrow \infty} \text{cum}\{2\tilde{\gamma}(h_1), \dots, 2\tilde{\gamma}(h_p)\} = 0, \quad (7)$$

$$h_j = \overline{0, n-1}, \quad j = \overline{1, p}, \quad p > 2, \quad n = 1, 2, 3, \dots$$

Доказательство. Используя свойства смешанных семиинвариантов из [2, с. 22], имеем

$$\begin{aligned} \text{cum}\{2\tilde{\gamma}(h_1), \dots, 2\tilde{\gamma}(h_p)\} &= \left[\prod_{j=1}^p (n - h_j) \right]^{-1} \times \\ &\times \sum_{s_1=1}^{n-h_1} \dots \sum_{s_p=1}^{n-h_p} \text{cum}\{(X(s_1 + h_1) - X(s_1))^2, \dots, (X(s_p + h_p) - X(s_p))^2\} = \\ &= \left[\prod_{j=1}^p (n - h_j) \right]^{-1} \sum_{s_1=1}^{n-h_1} \dots \sum_{s_p=1}^{n-h_p} \sum_{i_1, \dots, i_p=0}^2 m_{i_1} \dots m_{i_p} \times \\ &\times \text{cum}\left\{ X\left(s_1 + \left[\frac{i_1}{2}\right]h_1\right)X\left(s_1 + \left[\frac{i_1+1}{2}\right]h_1\right), \dots, X\left(s_p + \left[\frac{i_p}{2}\right]h_p\right)X\left(s_p + \left[\frac{i_p+1}{2}\right]h_p\right) \right\}, \end{aligned}$$

где

$$m_{i_j} = \begin{cases} 1, & i_j = 0; 2, \\ -2, & i_j = 1, \end{cases}$$

$j = \overline{1, p}$, $[i/2]$ – целая часть числа $i/2$.

Тогда на основании теоремы 2.3.2 [1], запишем

$$\begin{aligned} \text{cum}\{2\tilde{\gamma}(h_1), \dots, 2\tilde{\gamma}(h_p)\} &= \left[\prod_{j=1}^p (n - h_j) \right]^{-1} \sum_{\bigcup_{q=1}^M D_q = D} \sum_{s_1=1}^{n-h_1} \dots \sum_{s_p=1}^{n-h_p} \sum_{i_1, \dots, i_p=0}^2 m_{i_1} \dots m_{i_p} \times \\ &\times \prod_{q=1}^M \text{cum}\left\{ X\left(s_t + \left[\frac{i_t-1+r}{2}\right]h_t\right); (t, r) \in D_q \right\}, \end{aligned}$$

где суммирование ведется по всем неразложимым разбиениям D_q множества

$$D = \{(1, 1), (1, 2), (2, 1), (2, 2), \dots, (p, 1), (p, 2)\},$$

$$\text{cum}\left\{ X\left(s_t + \left[\frac{i_t-1+r}{2}\right]h_t\right); (t, r) \in D_q \right\}$$

означает смешанный семиинвариант от $X\left(s_t + \left[\frac{i_t-1+r}{2}\right]h_t\right)$ с индексами $(t, r) \in D_q$,

$$t = \overline{1, p}, \quad r = \overline{1, 2}, \quad \bigcup_{q=1}^M D_q = D, \quad M = \overline{1, 2p}.$$

Учитывая тот факт, что семинварианты порядка p , $p > 2$, гауссовских случайных процессов равны нулю, запишем

$$\text{cum}\{2\tilde{\gamma}(h_1), \dots, 2\tilde{\gamma}(h_p)\} = \left[\prod_{j=1}^p (n - h_j) \right]^{-1} \sum_{\bigcup_{q=1}^p D'_q = D} \sum_{s_1=1}^{n-h_1} \dots \sum_{s_p=1}^{n-h_p} \sum_{i_1, \dots, i_p=0}^2 m_{i_1} \dots m_{i_p} \times \\ \times \prod_{q=1}^p \text{cov} \left\{ X(s_t + \left[\frac{i_t - 1 + r}{2} \right] h_t); (t, r) \in D'_q \right\},$$

где суммирование ведется по всем неразложимым разбиениям $D'_q = \{(t_1, r_1), (t_2, r_2)\}$ множества $D = \{(1, 1), (1, 2), (2, 1), (2, 2), \dots, (p, 1), (p, 2)\}$, $t_1, t_2 = \overline{1, p}$, $r_1, r_2 = \overline{1, 2}$,

$$\text{cov} \left\{ X(s_t + \left[\frac{i_t - 1 + r}{2} \right] h_t); (t, r) \in D'_q \right\}$$

означает ковариацию $X(s_t + \left[\frac{i_t - 1 + r}{2} \right] h_t)$ с индексами $(t, r) \in D'_q$, $\bigcup_{q=1}^p D'_q = D$, $t = \overline{1, p}$, $r = \overline{1, 2}$.

Применяя связывающее соотношение

$$\gamma(h) = R(0) - R(h), \quad h \in Z,$$

между ковариационной функцией $R(h)$, $h \in Z$, и семивариограммой $\gamma(h)$, $h \in Z$, стационарного в широком смысле случайного процесса, учитывая (4), получаем требуемое предельное равенство (7).

Найдем асимптотическое распределение оценки $2\tilde{\gamma}(h)$, $h = \overline{0, n-1}$.

Теорема 4. При выполнении условия (4), оценка $2\tilde{\gamma}(h)$, задаваемая равенством (1), имеет асимптотически нормальное распределение с математическим ожиданием, равным $2\gamma(h)$, и предельной ковариационной структурой, удовлетворяющей соотношению (5).

Доказательство. Ранее показано, что $M\{2\tilde{\gamma}(h)\} = 2\gamma(h)$, $h = \overline{0, n-1}$, и ковариация построенной оценки (1) вариограммы удовлетворяет предельному равенству (5). Поскольку справедливо соотношение (7), то окончательное доказательство теоремы следует из теоремы 1.2 [2, с. 30].

ЛИТЕРАТУРА

1. Бриллинджер, Д. Временные ряды. Обработка данных и теория / Д. Бриллинджер. М., Мир, 1980. 536 с.
2. Труш, Н. Н. Асимптотические методы статистического анализа временных рядов / Н. Н. Труш. Мн, Белгосуниверситет, 1999. 218 с.
3. Труш, Н. Н. Исследование статистических свойств оценок вариограммы и ковариационной функции / Н. Н. Труш., Т. В. Цеховая // Вести НАН Беларуси. Сер. 1. 2001. № 2 - С. 24 – 29.
4. Цеховая, Т. В. Первые два момента оценки вариограммы гауссовского случайного процесса / Т. В. Цеховая // Дифференциальные уравнения и системы компьютерной алгебры. Материалы международной математической конференции. Брест, 2005. Ч. 2. С. 78 – 82.
5. Cressie, N. Fitting variogram models by weighted least squares / N. Cressie // Journal Inter. Assoc. Mathematical Geology. 1985. Vol. 17. № 5. P. 563 – 586.
6. Davis, B. M. Some Exact Sampling Distributions for Variogram Estimators / B. M. Davis, L. E. Borman // Journal Inter. Assoc. Mathematical Geology. 1979. Vol. 11; № 6. P. 643 – 653.
7. Цеховая, Т. В. Асимптотическое распределение оценки вариограммы / Т. В. Цеховая // Вестник БрГУ им. А.С. Пушкина. 2008. №2 (31). С. 32 – 37.