УДК 519.24

Т.В. Цеховая

АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ОЦЕНКИ ВАРИОГРАММЫ

Построена оценка вариограммы внутренне стационарного гауссовского случайного процесса с непрерывным 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.

В настоящее время проблема построения и изучения статистических свойств оценок основных характеристик временных рядов является достаточно актуальной [1–6]. 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 $X(s)$, $s\in R = (-\infty,+\infty)$, называется внутренне стационарным, если справедливы следующие равенства:

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

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

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

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

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

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

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

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

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

$h = \overline{0, n-1}$. 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:

$$M\{2\tilde{\gamma}(h)\} = 2\gamma(h),$$

$$\operatorname{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} =$$

$$= \frac{2}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} \{\gamma(t+h_1-s) + \gamma(t-s-h_2) - \gamma(t+h_1-s-h_2) - \gamma(t-s)\}^2, \quad (2)$$

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

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

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

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

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

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

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

$$\operatorname{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} =$$

$$= \frac{1}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} corr\{(X(t+h_1) - X(t))^2, (X(s+h_2) - X(s))^2\} \times$$

$$\times \sqrt{D(X(t+h_1) - X(t))^2 D(X(s+h_2) - X(s))^2} =$$

$$= \frac{2\{2\gamma(h_1)\}\{2\gamma(h_2)\}}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} \left\{ \frac{\gamma(t+h_1-s) + \gamma(t-s-h_2) - \gamma(t+h_1-s-h_2) - \gamma(t-s)}{\sqrt{2\gamma(h_1)}\sqrt{2\gamma(h_2)}} \right\}^2,$$

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,

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

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

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

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

то

$$\begin{gathered}\lim_{n\to\infty}(n-\min\{h_1,h_2\})\operatorname{cov}\{2\widetilde{\gamma}(h_1),2\widetilde{\gamma}(h_2)\} = \\ = 2\sum_{m=-\infty}^{+\infty}\{\gamma(m-h_2)+\gamma(m+h_1)-\gamma(m+h_1-h_2)-\gamma(m)\}^2,\end{gathered} \tag{5}$$

$$\lim_{n\to\infty}(n-h)D\{2\widetilde{\gamma}(h)\} = 2\Big[(2\gamma(h))^2 + 2\sum_{m=1}^{+\infty}\{\gamma(m-h)+\gamma(m+h)-2\gamma(m)\}^2\Big], \tag{6}$$

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

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

$$\operatorname{cov}\{2\widetilde{\gamma}(h_1), 2\widetilde{\gamma}(h_2)\} = \frac{2}{n-h_2}\Bigg[\sum_{m=-(n-h_2-1)}^{n-h_1-1}\{\gamma(m+h_1)+\gamma(m-h_2)-\gamma(m+h_1-h_2)-\gamma(m)\}^2 -$$

$$-\frac{2}{n-h_1}\sum_{m=1}^{n-h_1-1} m\{\gamma(m+h_1)+\gamma(m-h_2)-\gamma(m+h_1-h_2)-\gamma(m)\}^2\Bigg].$$

Аналогично рассуждаем для случая $h_1 < h_2$.

$$\operatorname{cov}\{2\widetilde{\gamma}(h_1), 2\widetilde{\gamma}(h_2)\} = \frac{2}{n-h_1}\Bigg[\sum_{m=-(n-h_2-1)}^{n-h_1-1}\{\gamma(m+h_1)+\gamma(m-h_2)-\gamma(m+h_1-h_2)-\gamma(m)\}^2 +$$

$$+\frac{2}{n-h_2}\sum_{m=-(n-h_2-1)}^{-1} m\{\gamma(m+h_1)+\gamma(m-h_2)-\gamma(m+h_1-h_2)-\gamma(m)\}^2\Bigg].$$

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

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

Следствие. Из теоремы 2 вытекает, что $\lim\limits_{n\to\infty} D\{2\widetilde{\gamma}(h)\} = 0$, $h = \overline{0, n-1}$.

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

Исследуем асимптотическое поведение семиинвариантов высших порядков статистики (1).

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

$$\lim_{n\to\infty} cum\{2\widetilde{\gamma}(h_1),\ldots, 2\widetilde{\gamma}(h_p)\} = 0, \tag{7}$$

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

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

$$cum\left\{2\tilde{\gamma}(h_1),\dots,2\tilde{\gamma}(h_p)\right\}=\left[\prod_{j=1}^{p}(n-h_j)\right]^{-1}\times$$

$$\times\sum_{s_1=1}^{n-h_1}\dots\sum_{s_p=1}^{n-h_p} cum\{(X(s_1+h_1)-X(s_1))^2,\dots,\ (X(s_p+h_p)-X(s_p))^2\}=$$

$$=\left[\prod_{j=1}^{p}(n-h_j)\right]^{-1}\sum_{s_1=1}^{n-h_1}\dots\sum_{s_p=1}^{n-h_p}\sum_{i_1,\dots,i_p=0}^{2} m_{i_1}\dots m_{i_p}\times$$

$$\times cum\left\{\ X(s_1+\left[\frac{i_1}{2}\right]h_1)X(s_1+\left[\frac{i_1+1}{2}\right]h_1),\dots,\ \ X(s_p+\left[\frac{i_p}{2}\right]h_p)X(s_p+\left[\frac{i_p+1}{2}\right]h_p)\ \right\},$$

где

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

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

Тогда на основании теоремы 2.3.2 [1] запишем

$$cum\left\{2\tilde{\gamma}(h_1),\dots,2\tilde{\gamma}(h_p)\right\}=\left[\prod_{j=1}^{p}(n-h_j)\right]^{-1}\sum_{\bigcup_{q=1}^{M}D_q=D}\ \sum_{s_1=1}^{n-h_1}\dots\sum_{s_p=1}^{n-h_p}\sum_{i_1,\dots,i_p=0}^{2} m_{i_1}\dots m_{i_p}\times$$

$$\times\prod_{q=1}^{M} cum\left\{\ X(s_t+\left[\frac{i_t-1+r}{2}\right]h_t);(t,r)\in D_q\ \right\},$$

где суммирование ведется по всем неразложимым разбиениям D_q множества

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

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

означает смешанный семиинвариант от

$$X(s_t+\left[\frac{i_t-1+r}{2}\right]h_t)$$

с индексами $(t,r)\in D_q$, $t=\overline{1,p}$, $r=\overline{1,2}$, $\bigcup_{q=1}^{M}D_q=D$, $M=\overline{1,2p}$.

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

$$cum\left\{2\tilde{\gamma}(h_1),\dots,2\tilde{\gamma}(h_p)\right\}=\left[\prod_{j=1}^{p}(n-h_j)\right]^{-1}\sum_{\bigcup_{q=1}^{p}D_q^{'}=D}\ \sum_{s_1=1}^{n-h_1}\dots\sum_{s_p=1}^{n-h_p}\sum_{i_1,\dots,i_p=0}^{2} m_{i_1}\dots m_{i_p}\times$$

$$\times\prod_{q=1}^{p}\operatorname{cov}\left\{\ X(s_t+\left[\frac{i_t-1+r}{2}\right]h_t);(t,r)\in D_q^{'}\ \right\},$$

где суммирование ведется по всем неразложимым разбиениям

$$D_q^{'}=\{(t_1,r_1),\ (t_2,r_2)\}$$

множества $D = \{(1,1),\ (1,2),\ (2,1),\ (2,2),\ldots,\ (p,1),\ (p,2)\}$, $t_1,t_2=\overline{1,p},\quad r_1,r_2=\overline{1,2},$

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

означает ковариацию

$$X(s_t+\left[\frac{i_t-1+r}{2}\right]h_t)$$

с индексами $(t,r)\in D_q^{'},\quad t=\overline{1,p},\quad r=\overline{1,2},\quad \bigcup_{q=1}^{p}D_q^{'}=D$.

Применяя связывающее соотношение

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

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

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

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 $M\{2\tilde{\gamma}(h)\}=2\gamma(h)$, $h=\overline{0,n\text{-}1}$ 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].

Исследования поддержаны Белорусским Республиканским Фондом фундаментальных исследований, грант № Ф07М-206.

СПИСОК ЛИТЕРАТУРЫ

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. – Минск : 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 : матер. междунар. математич. конф., 5–8 окт. 2005 г – 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. Borgman // Journal Inter. Assoc. Mathematical Geology. – 1979. – Vol. 11, № 6. – P. 643–653.

***T.V. Tsekhavaya*. Limiting Distribution of Variogram Estimator**

The paper deals with the problem of a statistical analysis of time series connected with the estimation of variogram. The limiting expressions of the first two moments and the higher order cumulants of the classical variogram estimator of Gaussian intrinsically stationary stochastic process with continuous time are presented. These expressions are then used to prove the theorem concerning the asymptotic distribution of the variogram estimator.