

УДК 519.24

СРАВНЕНИЕ ОЦЕНОК КОВАРИАЦИОННОЙ ФУНКЦИИ И СЕМИВАРИОГРАММЫ СЛУЧАЙНЫХ ПРОЦЕССОВ

Т. В. Цеховая

*Белорусский государственный университет, пр. Независимости, 4,
220030, г. Минск, Беларусь, Tsekhavaya@bsu.by*

Рассмотрены оценки ковариационной функции и семивариограммы стационарного случайного процесса и случайного процесса с линейным трендом. Выполнен сравнительный анализ построенных статистик.

Ключевые слова: случайный процесс; линейный тренд; оценка ковариационной функции, оценка семивариограммы.

COMPARISON OF THE ESTIMATORS OF COVARIANCE FUNCTION AND SEMIVARIOGRAM OF STOCHASTIC PROCESSES

T. V. Tsekhavaya

*Belarusian State University, Nezavisimosti Av., 4,
220030, Minsk, Belarus, Tsekhavaya@bsu.by*

Estimators of the covariance function and semivariogram of a stationary random process and a random process with a linear trend are considered. A comparative analysis of the constructed statistics is performed.

Keywords: stochastic process; linear trend; covariance function estimator; semivariogram estimator.

Семивариограмма и ковариационная функция являются основными мерами зависимости экспериментальных данных. Исследованию их свойств посвящена многочисленная литература [1-4]. Возникает задача построения оценок этих характеристик для различных случайных процессов и анализа свойств построенных статистик. Статистические свойства оценок ковариационных функций и семивариограмм стационарных случайных процессов исследовались, например, в [5, 6]. Данная статья является продолжением исследований [7] и посвящена построению и сравнению оценок ковариационной функции и семивариограммы для случайных процессов с линейным трендом.

Рассмотрим стационарный в широком смысле случайный процесс $Z(t), t \in T$, с неизвестными математическим ожиданием $E\{Z(t)\} = m_Z$, ковариационной функцией $R_Z(t), t \in T$, и семивариограммой $\gamma_Z(t), t \in T$.

Пусть $Z(1), Z(2), \dots, Z(n)$ – n последовательных, полученных через равные промежутки времени наблюдений за процессом $Z(t)$.

В качестве оценки ковариационной функции рассмотрим статистику

$$\bar{R}_Z(h) = \frac{1}{n-h} \sum_{s=1}^{n-h} (Z(s+h) - \bar{\mu}_Z)(Z(s) - \bar{\mu}_Z), \quad (1)$$

где $h = 0, \dots, n-1$, $\bar{\mu}_Z = \frac{1}{n} \sum_{i=1}^n Z(i)$, положим $\bar{R}_Z(-h) = \bar{R}_Z(h)$,

$h = 0, \dots, n-1$ и $\bar{R}_Z(h) = 0$ для $|h| \geq n$.

Статистические свойства оценки (1) ковариационной функции стационарного в широком смысле случайного процесса исследовались, например, в [7, 8]. Найдены выражения для моментов первых двух порядков построенной статистики. Показано, что если

$$\sum_{h=-\infty}^{+\infty} |R_Z(h)| < \infty, \quad (2)$$

то

$$\lim_{n \rightarrow \infty} E\{\bar{R}_Z(h)\} = R_Z(h)$$

и справедливо следующее соотношение:

$$|E\{\bar{R}_Z(h)\} - R_Z(h)| = O\left(\frac{1}{n}\right).$$

Заметим, что $\bar{R}_Z(h)$ является асимптотически несмещенной оценкой для ковариационной функции $R_Z(h)$.

Если имеют место условия (2) и

$$\sum_{h_1, h_2, h_3=-\infty}^{+\infty} |c_4(h_1, h_2, h_3)| < \infty, \quad (3)$$

где $c_4(h_1, h_2, h_3)$ – смешанный семиинвариант четвертого порядка процесса $Z(t)$, то для оценки $\bar{R}_Z(h)$, задаваемой равенством (1),

$$\lim_{n \rightarrow \infty} \text{cov}\{\bar{R}_Z(h_1), \bar{R}_Z(h_2)\} = 0, \quad h_1, h_2 = 0, \dots, n-1.$$

Отсюда вытекает, что

$$\lim_{n \rightarrow \infty} D\{\bar{R}_Z(h)\} = 0, \quad h = 0, \dots, n-1.$$

В качестве оценки семивариограммы случайного процесса $Z(t)$, $t \in T$, рассмотрим статистику вида

$$\bar{\gamma}_Z(h) = \frac{1}{2(n-h)} \sum_{s=1}^{n-h} (Z(s+h) - Z(s))^2, \quad (4)$$

где $h = 0, \dots, n-1$, положим $\bar{\gamma}_Z(-h) = \bar{\gamma}_Z(h)$, $h = 0, \dots, n-1$ и $\bar{\gamma}_Z(h) = 0$ для $|h| \geq n$.

В работе [7] были исследованы статистические свойства оценки семивариограммы (4) для стационарных в широком смысле случайных процессов. Показано, что статистика (4) является несмещенной оценкой семивариограммы, а если имеют место соотношения (2), (3), то

$$\lim_{n \rightarrow \infty} \text{cov}\{\bar{\gamma}_Z(h_1), \bar{\gamma}_Z(h_2)\} = 0,$$

$$\lim_{n \rightarrow \infty} D\{\bar{\gamma}_Z(h)\} = 0,$$

$h, h_1, h_2 = 0, \dots, n-1$. Таким образом, $\bar{\gamma}_Z(h)$ является также состоятельной в среднеквадратическом смысле оценкой для семивариограммы $\gamma_Z(h)$.

Отметим, что в работе (7) исследованы статистические свойства оценок (1) и (4) при дополнительных ограничениях на характеристики процесса $Z(t)$ как во временной, так и в частотной областях.

Таким образом, из анализа уже только первого момента построенных классических оценок семивариограммы $\bar{\gamma}_Z(h)$ и ковариационной функции $\bar{R}_Z(h)$ стационарных случайных процессов можно сделать вы-

вод, что удобнее использовать оценку семивариограммы (4), чем оценку ковариационной функции (1).

Далее рассмотрим случайный процесс с линейным трендом и построим оценки (1), (4) его характеристик. Сравнительный анализ указанных статистик позволит выявить наиболее предпочтительную оценку.

Пусть случайный процесс $Y(t), t \in T$ имеет вид:

$$Y(t) = X(t) + \varepsilon t + b, \quad \varepsilon, b \in R, \quad (5)$$

где $X(t), t \in T$ стационарный в широком смысле случайный процесс с математическим ожиданием $m_X = \text{const}$, ковариационной функцией $R_X(t)$, семивариограммой $\gamma_X(t), t \in T$.

Принимая во внимание (1), (4), построим оценки ковариационной функции $\bar{R}_Y(h)$ и семивариограммы $\bar{\gamma}_Y(h)$ для случайного процесса $Y(t)$. Получен следующий результат.

Теорема 1. Для оценки ковариационной функции $\bar{R}_Y(h)$ случайного процесса вида (5) имеет место следующее соотношение:

$$\bar{R}_Y(h) = \bar{R}_X(h) + G_{\bar{R}}(h),$$

где $h = 0, \dots, n-1$, $\bar{R}_X(h)$ - оценка (1) ковариационной функции стационарного процесса $X(t), t \in T$,

$$G_{\bar{R}}(h) = \frac{\varepsilon}{n-h} \sum_{s=1}^{n-h} \left(\left(s - \frac{n+1}{2} \right) (X(s+h) - \overline{\mu_X}) + \right. \\ \left. + \left(s+h - \frac{n+1}{2} \right) (X(s) - \overline{\mu_X}) \right) + \varepsilon^2 \left(\frac{n^2 - 2nh - 2h^2 - 1}{12} \right). \quad (6)$$

Доказательство. Принимая во внимание (5), запишем

$$\begin{aligned} \overline{\mu_Y} &= \frac{1}{n} \sum_{i=1}^n Y(i) = \frac{1}{n} \sum_{i=1}^n (X(i) + \varepsilon i + b) = \\ &= \frac{1}{n} \sum_{i=1}^n X(i) + \frac{1}{n} \sum_{i=1}^n \varepsilon i + \frac{1}{n} \sum_{i=1}^n b = \overline{\mu_X} + \varepsilon \frac{n+1}{2} + b. \end{aligned} \quad (7)$$

С учетом (1), (7) имеем

$$\begin{aligned}
\bar{R}_Y(h) &= \frac{1}{n-h} \sum_{s=1}^{n-h} (Y(s+h) - \overline{\mu_Y})(Y(s) - \overline{\mu_Y}) = \\
&= \frac{1}{n-h} \sum_{s=1}^{n-h} \left(X(s+h) + \varepsilon(s+h) + b - \overline{\mu_X} - \varepsilon \frac{n+1}{2} - b \right) \times \\
&\quad \times \left(X(s) + \varepsilon s + b - \overline{\mu_X} - \varepsilon \frac{n+1}{2} - b \right) = \\
&= \frac{1}{n-h} \sum_{s=1}^{n-h} \left(X(s+h) + \varepsilon \left(s+h - \frac{n+1}{2} \right) - \overline{\mu_X} \right) \times \\
&\quad \times \left(X(s) + \varepsilon \left(s - \frac{n+1}{2} \right) - \overline{\mu_X} \right).
\end{aligned}$$

Используя явный вид (1) оценки ковариационной функции случайного процесса $X(t)$, получим

$$\begin{aligned}
\bar{R}_Y(h) &= \bar{R}_X(h) + \frac{\varepsilon}{n-h} \sum_{s=1}^{n-h} \left(\left(s - \frac{n+1}{2} \right) (X(s+h) - \overline{\mu_X}) + \right. \\
&\quad \left. + \left(s+h - \frac{n+1}{2} \right) (X(s) - \overline{\mu_X}) + \varepsilon \left(s+h - \frac{n+1}{2} \right) \left(s - \frac{n+1}{2} \right) \right) = \\
&= \bar{R}_X(h) + \frac{\varepsilon}{n-h} \sum_{s=1}^{n-h} \left(\left(s - \frac{n+1}{2} \right) (X(s+h) - \overline{\mu_X}) + \right. \\
&\quad \left. + \left(s+h - \frac{n+1}{2} \right) (X(s) - \overline{\mu_X}) \right) + \varepsilon^2 \left(\frac{n^2 - 2nh - 2h^2 - 1}{12} \right),
\end{aligned}$$

что и требовалось доказать.

Теорема 2. Для оценки семивариограммы $\bar{\gamma}_Y(h)$ случайного процесса вида (5) имеет место следующее соотношение:

$$\bar{\gamma}_Y(h) = \bar{\gamma}_X(h) + G_{\bar{\gamma}}(h),$$

где $h = 0, \dots, n-1$, $\bar{\gamma}_Y(h)$ - оценка (4) семивариограммы стационарного процесса $X(t)$, $t \in T$,

$$G_{\bar{\gamma}}(h) = \frac{\varepsilon h}{(n-h)} \sum_{s=1}^{n-h} (X(s+h) - X(s)) + \frac{\varepsilon^2 h^2}{2}. \quad (8)$$

Доказательство. Учитывая (7), запишем выражение для оценки семивариограммы случайного процесса (5)

$$\begin{aligned} \bar{\gamma}_Y(h) &= \frac{1}{2(n-h)} \sum_{s=1}^{n-h} (Y(s+h) - Y(s))^2 = \\ &= \frac{1}{2(n-h)} \sum_{s=1}^{n-h} (X(s+h) + \varepsilon(s+h) - X(s) - \varepsilon s)^2. \end{aligned}$$

Принимая во внимание явный вид (4) оценки семивариограммы случайного процесса $X(t)$, имеем

$$\begin{aligned} \bar{\gamma}_Y(h) &= \bar{\gamma}_X(h) + \frac{1}{(n-h)} \sum_{s=1}^{n-h} (\varepsilon h X(s+h) - \varepsilon h X(s)) + \frac{1}{2(n-h)} \sum_{s=1}^{n-h} (\varepsilon^2 h^2) = \\ &= \bar{\gamma}_X(h) + \frac{\varepsilon h}{(n-h)} \sum_{s=1}^{n-h} (X(s+h) - X(s)) + \frac{\varepsilon^2 h^2}{2}, \end{aligned}$$

что и требовалось показать.

Исследуем поведение функций $G_{\bar{R}}(h)$ и $G_{\bar{\gamma}}(h)$ при $n \rightarrow \infty$ и фиксированных h .

Для этого смоделируем k , $k=8$ наборов наблюдений за процессом вида (5) с линейным трендом, $\varepsilon=0.2$, число наблюдений n_p , $p=\overline{1,8}$, в каждом наборе:

$$\begin{aligned} n_1 &= 50, n_2 = 100, n_3 = 150, n_4 = 200, \\ n_5 &= 300, n_6 = 400, n_7 = 500, n_8 = 600. \end{aligned}$$

В качестве $X(t)$, $t \in R$, рассмотрим стационарный случайный процесс с ковариационной функцией

$$R_X(t) = \frac{D}{1 + \alpha^2 t^2}, \quad D = 9, \alpha = 0,275,$$

семивариограммой [1]

$$\gamma_X(t) = R_X(0) - R_X(t) = 9 \left(1 - \frac{1}{1 + 0,275^2 t^2} \right), \quad t \in R$$

и смоделируем наблюдения за ним.

Применяя (6), (8), построим графики зависимостей функций $G_{\bar{R}}(h)$ и $G_{\bar{\gamma}}(h)$ от числа наблюдений n фиксированных h . Рассмотрим $h = 3$ и $h = 20$.

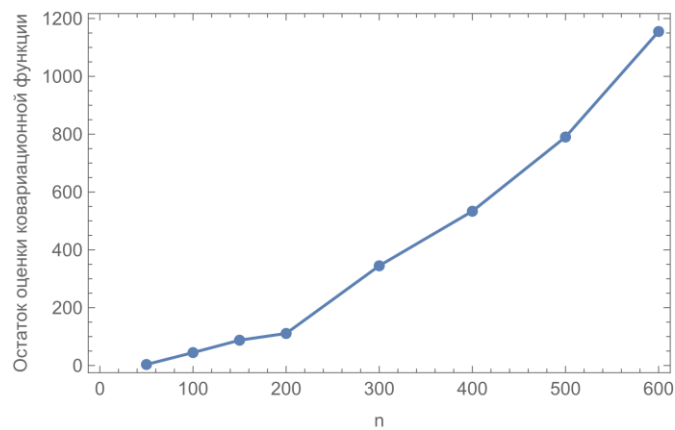


Рис. 1. $G_{\bar{R}}(h)$ при $h = 3$

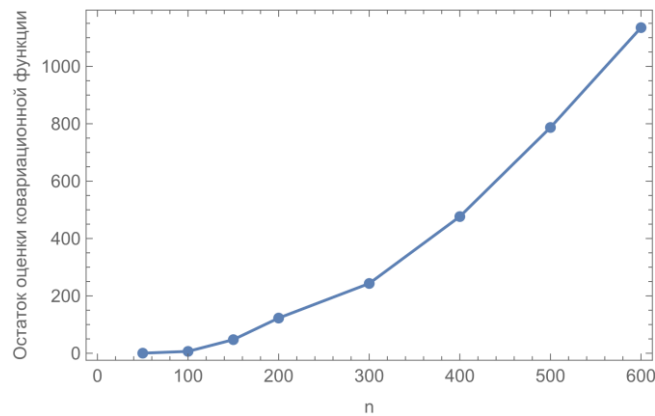


Рис. 2. $G_{\bar{R}}(h)$ при $h = 20$

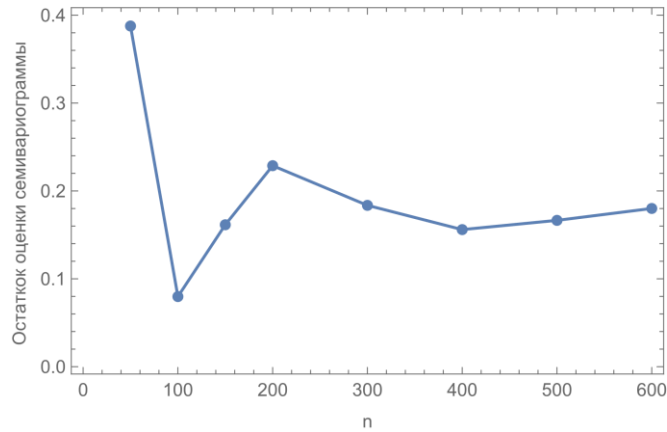


Рис. 3. $G_{\bar{\gamma}}(h)$ при $h=3$

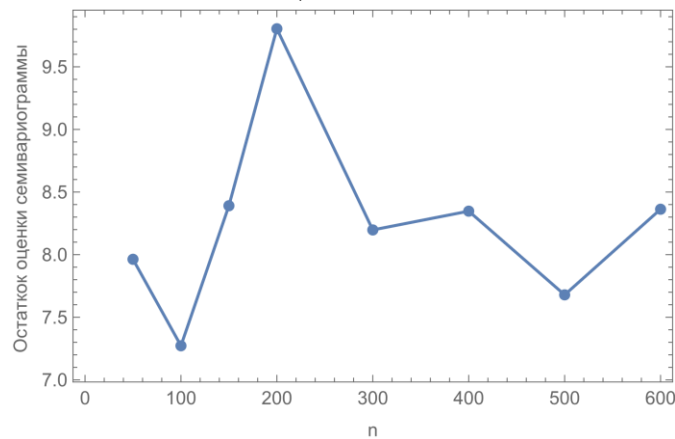


Рис. 4. $G_{\bar{\gamma}}(h)$ при $h=20$

Как видно из рисунков, функция $G_{\bar{R}}(h)$ при фиксированном h с ростом числа наблюдений возрастает, а функция $G_{\bar{\gamma}}(h)$ колеблется возле $\frac{1}{2}\varepsilon^2 h^2$. Заметим, что при $h=3$ значение $\frac{1}{2}\varepsilon^2 h^2 = 0.18$, а при $h=20$ $\frac{1}{2}\varepsilon^2 h^2 = 8$. Такое поведение функций говорит о большой стабильности оценки семивариограммы по сравнению с оценкой ковариационной функции в случае процесса с линейным трендом.

Таким образом, использование семивариограммы в анализе данных позволяет получать более устойчивые и надежные оценки зависимостей, что является значительным преимуществом в прикладных задачах. Это особенно важно в ситуациях, когда тренд невозможно удалить или его влияние необходимо учитывать при интерпретации результатов.

Библиографические ссылки

1. *Cressie N.* Statistics for Spatial Data. New York: Wiley Classics Library, 2015.
2. *Труш Н.Н., Цеховая Т.В.* Случайные процессы и их основные характеристики [Электронный ресурс] / Минск: БГУ. 2016. – ISBN 978-985-566-315-8. <https://elib.bsu.by/handle/123456789/158710>.
3. *Цеховая Т.В.* Свойства внутренне стационарных случайных процессов // Журн. Белорус. гос. ун-та. Математика. Информатика, 2017. № 1. С. 28–33.
4. *Цеховая Т. В.* Вариограммный анализ случайных процессов // Журн. Белорус. гос. ун-та. Математика. Информатика. 2017. № 2. С. 23–27.
5. *Андерсон Т.* Статистический анализ временных рядов. М.: Мир, 1976.
6. *Цеховая Т.В.* Асимптотическое распределение оценки семивариограммы гауссовского случайного процесса // Вестн.БГУ. Сер. 1, Физика. Математика. Информатика. 2015. № 1. С. 89–95.
7. *Цеховая Т.В., Труш Н.Н.* Оценки характеристик второго порядка во временной области стационарных процессов. Минск, БГУ. 2020. <https://elib.bsu.by/handle/123456789/254814>.
8. *Цеховая Т.В.* Первые два момента оценок семивариограммы и ковариационной функции // Наука и образование в современном мире: вызовы XXI века : Материалы III Международной научно-практической конференции. Том III. Нур-Султан, 2019 / Объединение юридических лиц в форме ассоциации "Общенациональное движение "Бобек"; редкол.: Х.Б. Маслов [и др.]. Астана, 2019. С. 199–204.