LIMITING DISTRIBUTION OF THE VARIOGRAM ESTIMATOR

T.V. Tsekhavaya, N.N. Troush
Belarus State University
Minsk, Belarus
e-mail: `Tsekhavaya@bsu.by`

Abstract

The paper deals with the problem of a statistical analysis of time series connected with the estimation of variogram. We present 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. These expressions are then used to prove the theorem concerning the asymptotic distribution of the variogram estimator.

1 Introduction

Consider a random process $X(s), s \in R$. Suppose further that $X(s), s \in R$, is a zero-mean, variance σ^2, intrinsically stationary Gaussian stochastic process with unknown covariance function $R(h), h \in R$, and variogram $2\gamma(h) = M(X(s+h)-X(s))^2, s, h \in R$, is given in Cressie (1985).

It is easy to show that $\{X(s+h)-X(s)\}^2 = 2\gamma(h)\cdot\chi_1^2$, where χ_1^2 denotes a chi-square random variable on 1 df. Therefore,

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

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

$$corr(\{X(t+h_1) - X(t)\}^2, \{X(s+h_2) - X(s)\}^2) =$$

$$\{corr(X(t+h_1) - X(t), X(s+h_2) - X(s))\}^2 =$$

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

where "corr" denotes correlation.

The variogram estimator $2\tilde{\gamma}(h)$ in terms of sequence of observations $X(1), ..., X(n)$, is defined as

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

$h = \overline{0, n-1}$, with $2\tilde{\gamma}(-h) = 2\tilde{\gamma}(h), h = \overline{0, n-1}$ and $2\tilde{\gamma}(h) = 0$, for $\mid h \mid\geq n$.

It is the purpose of this paper to derive the asymptotic distribution of the variogram estimator $2\tilde{\gamma}(h), h = \overline{0, n-1}$. The approach is similar to the approach taken in the time series literature, and the reader is referred to Brillinger (1975), Troush (1999), Tsekhavaya (2002) for theorems regarding the asymptotic distribution of the spectral density estimator, covariance estimator and variogram estimator of a time series.

2 First-, Second-order Moments

In order to state and prove the theorem concerning the asymptotic distribution of the variogram estimator $2\tilde{\gamma}(h), h = \overline{0, n-1}$, it is first necessary to calculate the first two moments of the examined estimator (2).

The classical variogram estimator $2\tilde{\gamma}(h)$ is unbiased for $2\gamma(h)$.

Theorem 1. *For the variogram estimator* $2\tilde{\gamma}(h)$, $h = \overline{0, n-1}$, *the expressions*

$$cov\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} = \tag{3}$$

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

$$D\{2\tilde{\gamma}(h)\} = \frac{2}{(n-h)^2} \sum_{t,s=1}^{n-h} \{\gamma(t-s+h) + \gamma(t-s-h) - 2\gamma(t-s)\}^2 \tag{4}$$

are valid, where $h, h_1, h_2 = \overline{0, n-1}$.

Proof. From the definition of the covariance, we can show that $cov\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} =$

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

Using the relation (1)

$$cov\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} = \frac{2\{2\gamma(h_1)\}\{2\gamma(h_2)\}}{(n-h_1)(n-h_2)} \times$$

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

and hence (3) is valid. Introducing $h_1 = h_2 = h$, it is easy to show (4).

We shall found the limiting expressions of the second-order moments of the variogram estimator (2).

Theorem 2. *Let*

$$\textstyle\sum_{r=-\infty}^{+\infty} \mid R(r) \mid < \infty. \tag{5}$$

Then

$$\lim_{n\to\infty} (n - min\{h_1, h_2\}) cov\{2\gamma(h_1), 2\gamma(h_2)\} = \tag{6}$$

$$= 2 \sum_{m=-\infty}^{+\infty} \{\gamma(m-h_2) + \gamma(m+h_1) - \gamma(m+h_1-h_2) - \gamma(m)\}^2,$$

$$\lim_{n\to\infty} (n-h) D\{2\gamma(h)\} = 2\{(2\gamma(h))^2 + 2\sum_{m=1}^{+\infty} (\gamma(m-h) + \gamma(m+h) - 2\gamma(m))^2\}, \tag{7}$$

where $h, h_1, h_2 = \overline{0, n-1}$.

Proof. Consider (3). Assume that $h_1 > h_2$. Introducing $t = t, t - s = m$, we write

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

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

A similar expression can be derived for $h_1 < h_2$. Hence, applying (5), we obtain (6).

The limiting expression (7) for variance of the classical variogram estimator $2\gamma(h)$ follows immediately from (6) if $h_1 = h_2 = h$.

Corollary 1. *Let all the assumptions of Theorem 2 be satisfied. Then it follows that* $\lim_{n\to\infty} cov\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} = 0, \quad \lim_{n\to\infty} D\{2\tilde{\gamma}(h)\} = 0,\ h, h_1, h_2 = \overline{0, n-1}$.

3 Higher Order Cumulants

In order to found the asymptotic distribution of the variogram estimator $2\tilde{\gamma}(h)$ it is necessary to investigate an asymptotic behavior of the cumulant $cum\{2\tilde{\gamma}(h_1), ..., 2\tilde{\gamma}(h_p)\}$, $h_j = \overline{0, n-1}, j = \overline{1, p}$.

Theorem 3. *The cumulant of the variogram estimator (2) is*

$$cum\{2\tilde{\gamma}(h_1), ..., 2\tilde{\gamma}(h_p)\} = \{\prod_{j=1}^{p}(n-h_j)\}^{-1}\times \tag{8}$$

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

where $h_j = \overline{0, n-1}, j = \overline{1, p}$, *the summation over* $D = \bigcup_{q=1}^{p} D'_q$ *is over all indecomposable partitions* $D'_q = \{(t_1, r_1), (t_2, r_2)\}$, $t_1, t_2 = \overline{1, p}$, $r_1, r_2 = \overline{1, 2}$, *of the set* $D = \{(1,1), (1,2), (2,1), (2,2), ..., (p,1), (p,2)\}$, $cov\{X(s_t + [\frac{i_t-1+r}{2}]h_t); (t,r) \in D'_q\}$ *is the covariance of* $X(s_t + [\frac{i_t-1+r}{2}]h_t)$ *with* $(t,r) \in D'_q$, $t = \overline{1,p}, r = \overline{1,2}$, $[\frac{i}{2}]$ *is the whole part of number* $\frac{i}{2}$, $m_{i_j} = \{1,\ if\ i_j = 0, 2; -2,\ if\ i_j = 1.\}$

Proof. Applying the properties of a sample cumulants Brillinger (1975), we obtain

$$cum\{2\tilde{\gamma}(h_1), ..., 2\tilde{\gamma}(h_p)\} = \{\prod_{j=1}^{p}(n-h_j)\}^{-1}\times$$

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

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

$$\times cum\{X(s_1+[\tfrac{i_1}{2}]h_1)X(s_1+[\tfrac{i_1+1}{2}]h_1),...,X(s_p+[\tfrac{i_p}{2}]h_p)X(s_p+[\tfrac{i_p+1}{2}]h_p)\},$$

Let $Y_t = X(s_t+[\frac{i_t}{2}]h_t)X(s_t+[\frac{i_t+1}{2}]h_t), t=\overline{1,p}$. Lemma 2.5 of Troush (1999) then yields

$$cum\{2\tilde{\gamma}(h_1),...,2\tilde{\gamma}(h_p)\} = \{\textstyle\prod_{j=1}^{p}(n-h_j)\}^{-1}\times$$

$$\times \sum_{D=\bigcup_{q=1}^{M} D_q} \sum_{s_1=1}^{n-h_1} ... \sum_{s_p=1}^{n-h_p} \sum_{i_1,...,i_p=0}^{2} m_{i_1}...m_{i_p} \prod_{q=1}^{M} cum\{X(s_t+[\frac{i_t-1+r}{2}]h_t);(t,r)\in D_q\},$$

where the summation over $D=\bigcup_{q=1}^{M} D_q$ is over all indecomposable partitions of the set D, $cum\{X(s_t+[\frac{i_t-1+r}{2}]h_t);(t,r)\in D_q\}$, is the cumulants of $X(s_t+[\frac{i_t-1+r}{2}]h_t)$ with $(t,r)\in D_q,\ t=\overline{1,p},\ r=\overline{1,2},\ \bigcup_{q=1}^{M} D_q = D,\ M=\overline{1,2p}$.

Because $MX(s)=0$ and the sample cumulants of the Gaussian stationary stochastic process $cum\{X(t_1),...,X(t_{p-1})\}=0, p>2$, then (8) is valid.

Theorem 4. *Let all the assumptions of Theorem 3 be satisfied. Then*

$$\lim_{n\to\infty} cum\{2\tilde{\gamma}(h_1),...,2\tilde{\gamma}(h_p)\} = 0, \tag{9}$$

where $2\tilde{\gamma}(h)$ *is the variogram estimator given by (2),* $h_j=\overline{0,n-1}, j=\overline{1,p}, p\geq 2$.

Proof. The result follows directly from (5).

Theorem 5. *Let* $\sum_{h=-\infty}^{+\infty}|\gamma(h)|<\infty$. *The conclusion (9) of Theorem 4 then holds.*

Proof. The proof of this theorem follows directly from the relation between the covariance function and variogram of a zero-mean second-order-stationary stochastic process:

$$2\gamma(h) = 2\{R(0)-R(h)\}, h\in R.$$

4 Asymptotic Distribution

Theorem 6. *Let all the assumptions of Theorem 4 or Theorem 5 be satisfied. Then the variogram estimator (2) is asymptotically normally distributed with mean* $2\gamma(h), h\in R$, *and asymptotic variance (7).*

Proof. As we have mentioned before, $M\{2\tilde{\gamma}(h)\}=2\gamma(h)$, and the conditions (7) and (9) are valid. From Theorem 1.2 of Troush (1999), we have the final proof of Theorem.

References

[1] Brillinger D.R. (1975). *Time Series, Data Analysis and Theory*. Holt, New York.

[2] Cressie N. (1985). Fitting variogram models by weighted least squares. *Journal Inter. Assoc. Math. Geol.* Vol. **17(5)**, pp. 563-586.

[3] Troush N.N. (1999). *Asymptotic Methods of Statistical Analysis of Time Seriesy.* Minsk, BSU. (in Russian).

[4] Tsekhavaya T.V. (2002). Asymptotic distribution of the variogram estimator of the stationary stochastic process. *Bulletin BSU* . Vol. **1(1)**, pp. 104-105. (in Russian).