On Estimating Statistical Characteristics of Red-noise Processes

M I Fortus,  
A. M. Obukhov Institute of Atmospheric Physics RAS, Pyzhevsky per. 3, Moscow, 119017, Russia  
Corresponding author: M I Fortus mfortus@yandex.ru

Abstract. By now the problem of estimating statistical characteristics of red-noise processes is not completely developed. However this problem is of considerable importance for statistical climatology because in the recent years extensive literature has been published containing empirical spectra of red-noise type having noticeable regions of a monotonic increase when a time-scale \( T \) increases (with a frequency \( \omega \) decrease). Such spectra are commonly approximated by the functions increasing like \( \omega^{-2\alpha}, 0 < \alpha < 1/2 \) (red-noise spectra). When dealing with a problem of estimating statistical characteristics of red-noise processes, the estimate errors are to be calculated. As a rule estimate errors become acceptable small when the empirical time series is long enough and asymptotic formulas for the errors can be used. In this paper, we give a list of formulas describing asymptotic behavior of the estimate errors for mean values, covariances, trend parameters, and structure functions for a wide interval of values of \( \alpha \).

Introduction

When analyzing time series obtained both from observations and from numerical modeling, it is common in statistical climatology to use methods and terminology of the random functions theory, namely, the theory of statistical estimation. The assumption that the available time series is a segment of the realization of a stationary random process, allows to estimate its statistical characteristics: a mean value \( m \), a variance \( \sigma^2 \), a covariance function \( B(T) \), etc. However, any sample estimate is informative if it is accompanied with an estimation error or a confidence interval. It is common to assume that the covariance function tends to zero so rapidly that it is integrable (i.e. the spectral density \( f(\omega) \) is bounded). Then mean square estimation errors tend to zero as a rate of \( \text{const} / \sqrt{T} \) or faster when \( T \to \infty \) [1]. Just this rate of decrease is used in many of algorithms of statistical processing of time series when calculating estimate errors.

As a rule, estimate errors become acceptable small when \( T \) is large enough so that asymptotic formulas for the errors can be used. However, for a very large \( T \) the covariance function \( B(T) \) cannot be estimated from observations and it is necessary to make additional hypotheses for its behavior when \( T \to \infty \) (like that mentioned above). Some papers propose to approximate the asymptotic behavior of \( B(T) \) by the power function \( \text{const} T^{2\alpha-1}, 0 < \alpha < 1/2 \). The random process with such a covariance function is stationary though \( B(T) \) is nonintegrable and the corresponding spectral density \( f(\omega) \) tends...
to infinity with $\omega \to 0$ [1]. In this case the estimate errors (and the width of confidence intervals as well) tend to zero slower than $\text{const} / \sqrt{T}$. A statistically stationary process with the above-mentioned $B(T)$ and $f(\omega)$ is called a red-noise process.

In this paper, we give without proofs the summary list of formulas for the estimate errors for characteristics the most used in geophysics: mean values, covariances, trend parameters, and structure functions. The results for a bounded spectral density ($\alpha = 0$) are well known [1] (at least for mean values and variances) but the most of results for the case $0 < \alpha < 1/2$ are obtained by the author. The calculations have been performed for continuous varying argument $t$. The corresponding results for a discrete argument do not differ essentially from the continuous case.

1. Results

Let

$$x(t), -T/2 < t < T/2$$

be a realization of a red-noise process with the spectral density

$$f(\omega) = g(|\omega|)/|\omega|^{2\alpha}, 0 < \alpha < 1/2$$

(1)

where $g(|\omega|) \geq 0, g(0) > 0, g(\omega) < G < \infty$, and $\lim(g(|\omega|)) = 0$ when $\omega \to \infty$ so rapidly that $f(\omega)$ is integrable.

The asymptotics for the corresponding covariance function

$$B(T) \sim \alpha(2\alpha + 1)\sigma^2 T^{2\alpha - 1}[2g(0)\Gamma(1 - 2\alpha)\sin \pi \alpha], T \to \infty$$

holds for a wide class of processes.

1.1. Estimation of the mean value

The mean square estimate is

$$m^* = \overline{x}, -T/2 < t < T/2$$

(2)

(The overbar implies averaging over $t$). The symmetric about zero interval was chosen in order to simplify calculations. This estimate is unbiased, so the estimate error is equal to the standard deviation of $m^*$:

$$\delta_T(m^*) = \sigma_T(m^*).$$

The asymptotic behaviour of the error squared $\delta_T^2(m^*)$ is

$$\delta_T^2(m^*) \sim T^{2\alpha - 1} \frac{2g(0)\Gamma(1 - 2\alpha)\sin \pi \alpha}{\alpha(1 + 2\alpha)}, T \to \infty, -1/2 < \alpha < 1/2.$$  

(3)

(For simplicity sake here and below we give asymptotics for the characteristics squared). Noteworthy is that this asymptotics holds for negative values of $\alpha$ as well).

The problems related to mean value estimating were investigated in detail in [2].

1.2. Estimation of the variance

$$\delta_T(\sigma^2*) = \delta_T(B(0)^*) = E(B(0)^* - B(0)).$$

(4)

a). The mean value $m$ is known

$$B(0)^* = \sigma^{2*} = (x(t) - m)^2.$$  

(5)

In this case the estimate $\sigma^{2*}$ is nonbiased. In addition, suppose that $x(t)$ is normally distributed.

If $1/4 < \alpha < 1/2$
\[ \delta_T^2 (\sigma^2) \sim T^{4\alpha - 2} \frac{4g^2(0)\Gamma^2 (1-2\alpha)\sin^2 \pi \alpha}{\alpha (4\alpha - 1)}, 1/4 < \alpha < 1/2, T \to \infty. \]  

If \(-1/2 < \alpha < 1/4\)

\[ \delta_T^2 (\sigma^2) \sim (2\pi/T) h(0), T \to \infty. \]

Here

\[ h(\omega) = 2 \int_{-\infty}^{\infty} \frac{g(\omega - \mu)g(\mu)}{|\omega - \mu|^{1/2} |\mu|^{1/2}} d\mu = O\left(\ln \omega\right), \omega \to \infty. \]

If \(\alpha = 1/4\)

\[ \delta_T^2 (\sigma^2) \sim (2/T^2) \int_{-\infty}^{\infty} h(\omega) \frac{(1 - \cos \omega T)}{\omega^2} d\omega = O\left((\ln T)/T\right), T \to \infty. \]

**b). The mean value \(m\) is unknown**

\[ \sigma^2 = B(0)^* = (\bar{x}(t) - \bar{x})^2. \]

In this case the bias of \(\sigma^2\) is negative:

\[ E\{\sigma^2\} = \sigma^2 - \sigma_T^2\left(m^*\right). \]

So the estimate error squared is

\[ \Delta_T^2 (\sigma^2) = \delta_T^2 (\sigma^2) + \sigma_T^4\left(m^*\right). \]

After cumbersome calculations we get

**If \(1/4 < \alpha < 1/2\)**

\[ \Delta_T^2 (\sigma^2) \sim T^{4\alpha - 2} \frac{4g^2(0)\Gamma^2 (1-2\alpha)\sin^2 \pi \alpha}{\alpha (4\alpha - 1)} \left(1 + \frac{1}{\alpha(1+2\alpha)}\right), T \to \infty. \]

It is seen that if \(1/4 < \alpha < 1/2\) and \(m\) is unknown the rate of decrease of the estimate error squared is the same as in the case of \(m\) known (6) but the coefficient is somewhat more at the cost of the second summand in the parenthesis.

**If \(-1/2 < \alpha < 1/4, \alpha \neq 0\)**

\[ \Delta_T^2 (\sigma^2) \sim 2\pi h(0)/T, T \to \infty. \]

**If \(\alpha = 0\)**

\[ \Delta_T^2 (\sigma^2) \sim \text{const}/T, T \to \infty \]

[1], page228.

**1.3. Estimation of the coefficients of a linear trend**

Suppose that the process \(y(t)\) is a sum of a linear function and a random stationary process with zero mean and the spectral density (1):

\[ y(t) = a + bt + x(t), -T/2 < t < T/2. \]

Mean square linear estimates of coefficients \(a\) and \(b\) are:

\[ a^* = \bar{y}, \]

\[ b^* = \bar{y}/t^2. \]

The both estimates are nonbiased:

\[ E\{a^*\} = a, E\{b^*\} = b. \]
So, their mean square errors are equal to their standard deviations, \( \sigma_y (a^*) \) and \( \sigma_y (b^*) \). The interval for \( t \) is symmetrical about zero (15) – so the mean value of \( y(t) \) is \( a \) and all the formulas for \( m^* \) can be used for \( a^* \) (including the asymptotic behavior (3)).

**Asymptotics of the variance \( \sigma_y^2 (b^*) \) (which is equal to \( \delta_y^2 (b^*) \))**

We have:

\[
\sigma_y^2 (b^*) = E \{ (b^* - b)^2 \} = \left( 1/T^2 \right) \int_{-\infty}^{\infty} ds B(t-s) ds.
\]

Using the equality

\[
t^2 = T^2 / 12 ,
\]

it is not difficult to obtain

\[
\sigma_y^2 (b^*) \sim T^{2 \alpha - 3} \frac{72 g(0) \Gamma(2 - 2 \alpha) \sin \pi \alpha}{\alpha (2 \alpha + 1)(2 \alpha + 3)}, -1/2 < \alpha < 1/2 .
\] (18)

Note that \( \sigma_y^2 (b^*) \) tends to zero faster than \( \sigma_y^2 (a^*) \). This can be explained by the fact that the estimate \( b^* \) is expressed via differences \( (x(t) - x(-t)) \) whose spectral density tends to zero when \( \omega \to 0 \).

**Approximation of the realization \( y(t), -T/2 < t < T/2 \), by the linear trend \( a + bt \)**

and **asymptotical behavior of the error \( d(T) \) of this approximation.**

The error squared \( d^2 (T) \) can be expressed via \( \sigma_y^2 (a^*) \) and \( \sigma_y^2 (b^*) \):

\[
d^2 (T) = B(0) - \sigma_y^2 (a^*) - t^2 \sigma_y^2 (b^*). \] (19)

When \( T \to \infty \) we have

\[
d^2 (T) \sim B(0) - T^{2 \alpha - 1} \frac{(12 - 8 \alpha) g(0) \Gamma(1 - 2 \alpha) \sin \pi \alpha}{\alpha (2 \alpha + 1)(2 \alpha + 3)}, -1/2 < \alpha < 1/2 . \] (20)

It is seen that \( d^2 (T) \) increases when \( T \to \infty \) and approaches \( B(0) \). It is not surprising because the longer is the realization, the worse it can be approximated by a linear trend which is determined by but two parameters, \( a \) and \( b \), and cannot describe a broad spectrum of fluctuations.

**1.4. The processes with stationary increments**

Let the time series

\[
y(t), -T/2 < t < T/2
\]

is a realization of a process with stationary increments \( y(t) \) with the mean value

\[
E \{ y(t) \} = c_o + c_1 t .
\]

and the structure function

\[
D(\tau) = E \{ [y'(t + \tau) - y'(t)]^2 \}, y'(t) = y(t) - E \{ y(t) \} .
\]

Its spectral representation is

\[
D(\tau) = 4 \int_0^\infty (1 - \cos \omega \tau) f_1(\omega) d\omega
\] (21)

[1].
By analogy with the spectral density class (1) considered above, suppose that in the vicinity of zero \( f_i(\omega) \) has a form
\[
f_i(\omega) = g_i(\omega) / |\omega|^{2\alpha+2}, \quad -1/2 < \alpha < 1/2
\]
where \( g_i(\omega) \) is an even nonnegative function for all \( \omega \) with \( g_i(0) > 0 \) and \( g_i(\omega) < G_i < \infty \), and \( f_i(\omega) \) is integrable.

For every fixed \( \tau \) the time series generated by the increments
\[
z_\tau(t) = y(t + \tau) - y(t)
\]
can be considered as a realization of a stationary process (depending of \( \tau \) parametrically). The mean value \( E\{z_\tau(t)\} \) equals to \( c_1 \tau \) and the spectral density of \( z_\tau(t) \) is
\[
f_2(\omega) = 2(1 - \cos \omega \tau) f_1(\omega) = 2(1 - \cos \omega \tau) g_1(\omega) / |\omega|^2.
\]
(22)

It follows from the suppositions made above that this spectral density belongs to the class (1) if we set
\[
g(\omega) = g_2(\omega) = 2(1 - \cos \omega \tau) g_1(\omega) / |\omega|^2.
\]
so that
\[
g(0) = g_2(0) = \tau^2 g_1(0).
\]

It means that we can use the results obtained above. Here we dwell on

**Estimation of the constant** \( c_1 \).

The estimate
\[
c^*_1 = z_\tau(t) / \tau
\]
is unbiased and the estimate error squared equals to its variance \( \sigma^2_\tau(c^*_1) \).

All the results of p.2.1 relating to the estimate of the mean value \( m^* \) are valid for \( c^*_1 \) as well (if \( f(\omega) \) and \( \sigma^2_z(m) \) are replaced by \( f_1(\omega) \) and \( \tau^2 \sigma^2_z(c^*_1) \) correspondingly).

**Asymptotics of the variance of** \( c^*_1 \)

The formula (3) turns to
\[
\sigma^2_\tau(c^*_1) \sim (1/T^{1-2\alpha}) \left[ 2 g(z_1(0)) \Gamma(1-2\alpha) \sin \pi \alpha / \alpha (1+2\alpha) \right], \quad -1/2 < \alpha < 1/2.
\]
(23)

In particular,
\[
\sigma^2_\tau(c^*_1) \sim 2\pi g_1(0) / T (\alpha = 0).
\]

**Estimation of the structure function**

The structure function \( D(t) \) is a variance of \( z_\tau(t) \). If \( c_1 \) is known the estimate \( D^*(\tau) \) is nonbiased and the estimate error squared equals to the variance of \( D^*(\tau) \).

**Asymptotics of the variance of** \( D^*(\tau) \)

The asymptotics in question is
\[
\delta^2_\tau(D^*(\tau)) \sim (1/T^{2-4\alpha}) \frac{4 g^2_2(0) \Gamma^2(1-2\alpha) \sin^2 \pi \alpha}{\alpha (4\alpha - 1)}, \quad 1/4 < \alpha < 1/2,
\]
\[
\delta^2_\tau(D^*(\tau)) \sim 2\pi h_2(0) / T, \quad -1/2 < \alpha < 1/4,
\]
\[
\delta^2_\tau(D^*(\tau)) \sim O\{\ln T / T\}, \quad \alpha = 1/4
\]
(as in (7-9)). Here \( h_2(0) \) is determined via \( f_2(\tau) \) by the formula analogous to (22).
Conclusions
In this study, we have obtained formulas for an asymptotic behavior of the estimate errors for the main statistical characteristics of stationary random processes and processes with stationary increments. Different forms of the densities were considered including those of “red-noise” processes with their nonbounded spectral densities in the vicinity of zero. We have shown that for the processes just mentioned, the errors’ asymptotics is slower than for processes whose spectral densities are bounded everywhere. This property of estimate errors in the red-noise case leads to increase of confidence intervals which fact should be taken into account when estimating statistical characteristics. In addition, the problem of estimating trend parameters and structure function is considered.

References
[1] Yaglom A M 1987 Correlation Theory of Stationary and Related Functions I Basic Results (New York Springer Verlag)
[2] Fortus M I 2010 Izv. Atmos. Ocean. Phys. 46 563