Russian Physics Journal, Vol. 40, No. 4, 1997

TESTING OF VARIANCE HYPOTHESES FOR NONSTATIONARY WHITE GAUSSIAN NOISE

E. E. Zmeeva and A. F. Terpugov

UDC 519.2:53.05

Testing of hypotheses as to the equality and proportionality of noise variance to specified time functions is discussed.

Analysis of noise in radio communications channels is of very great importance in the design of HF and UHF radio communication systems.

White Gaussian noise, in which noise measurements x_i made at certain times are normally distributed independent random variables with zero mean value $M(x_i) = 0$, offers a fairly good mathematical model of noise of this kind. White Gaussian noise is usually understood as noise in which the variances of all x_i are equal, i.e., $D(x_i) = \sigma^2$. However, the noise encountered in practice is usually not stationary, i.e., the variances $D(x_i) = \sigma^2_i$ vary in time. We shall henceforth refer to such a process as nonstationary white noise.

Let the noise measurements be made at equal time intervals, and let $\{x_1, x_2, \ldots, x_N\}$ be the measured values. Below we study a mathematical model of the noise in which the joint probability density of the $\{x_i\}$ has the form

$$p(x_1, x_2, \ldots x_N) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma_i^2}\right), \tag{1}$$

and refer to it as nonstationary white Gaussian noise.

Many problems arise in study of such processes. In this paper we shall discuss problems associated with testing of hypotheses as to the behavior of the noise variance σ^2_i as a function of i, i.e., actually as a function of time.

TESTING THE HYPOTHESIS THAT THE VARIANCE IS EQUAL TO A GIVEN FUNCTION

Consider testing of a statistical hypothesis of the form:

$$\begin{gathered} H_0 : D\{x_i\} = f(i),\ i = \overline{1, N}, \\ H_1 : D\{x_i\} \neq f(i). \end{gathered} \tag{2}$$

In other words, test whether the dependence of $D\{x_i\}$ on i has the specified form f(i).

To test this hypothesis, we construct a statistic that embodies the following basic idea: find a statistic S whose mean value is equal to zero under hypothesis H_0 but always greater than zero under the alternative H_1. It is desirable that the variance of this statistic be equal to 1 at least in the asymptotic case $N \to \infty$. Then the following decision rule will appear quite natural: if $S < C$, where C is a certain threshold constant, adopt hypothesis H_0, but if $S \geq C$, adopt the alternative H_1.

Let us describe the basic steps in construction of this statistic. Since the hypothesis H_0:$D\{x_i\} = f(i)$, is to be verified, it is natural to take a statistic of the form

Anzhero-Sudzhensk Branch, Tomsk Pedagogical University. Tomsk State University. Translated from Izvestiya Vysshikh Uchebnykh Zavedenii, Fizika, No. 4, pp. 56-62, April, 1997. Original article submitted July 2, 1996.

1064-8887/97/4004-0363$18.00

$$S_1=\sum_{i=1}^{N}(x_i^2-f(i))^2. \tag{3}$$

Let the true variance be $D\{x_i\} = f_1(i)$, where $f_1(i)$ is a certain arbitrary function. We find the mean value of statistic S_1. Expanding the square and recognizing that $M\{x_i^2\} = f_1(i)$, $M(x_i^4\} = 3f_1^2(i)$ for Gaussian random variables, we obtain

$$M\{S_1|H_1\}=\sum_{i=1}^{N}(f_1(i)-f(i))^2+2\sum_{i=1}^{N}f_1^2(i). \tag{4}$$

We see that the first term in (4) has the required property, since it vanishes when $f_1(i) = f(i)$ and is always positive when $f_1(i) \neq f(i)$. However, the second term in (4) is redundant and must be removed. Since this term owes its appearance to the coefficient 3 in $M\{x^4_i\}$, it will disappear if this coefficient becomes equal to 1. Therefore the corrected statistic has the form

$$S_2=\frac{1}{3}\sum_{i=1}^{N}x_i^4-2\sum_{i=1}^{N}x_i^2f(i)+\sum_{i=1}^{N}f^2(i). \tag{5}$$

Now evaluation of the mean value of this statistic gives

$$M\{S_2|H_1\}=\sum_{i=1}^{N}(f_1(i)-f(i))^2, \tag{6}$$

from which we see that $M\{S_2|H_0\} = 0$; $M\{S_2|H_1\} > 0$, i.e., this statistic meets the required condition.

To obtain the final form of this statistic we calculate its variance under hypothesis H_0, when $D\{x_i\} = f(i)$. Remembering that the x_i are independent, we obtain

$$D\{S_2|H_0\}=\frac{8}{3}\sum_{i=1}^{N}f^4(i). \tag{7}$$

For practical application it is advantageous to normalize S_2 in such a way that its variance will be 1 under hypothesis H_0. Therefore the final form of statistic S in testing the hypothesis that the variance is equal to the given function f(i) should be

$$S=\sum_{i=1}^{N}\left(\frac{x_i^4}{3}-2x_i^2f(i)+f^2(i)\right)\Bigg/\sqrt{\frac{8}{3}\sum_{i=1}^{N}f^4(i)}. \tag{8}$$

The following properties will be satisfied for it:

$$\begin{gathered}M\{S|H_0\}=0;\ D\{S|H_0\}=1;\\ M\{S|H_1\}=\sum_{i=1}^{N}(f_1(i)-f(i))^2\Bigg/\sqrt{\frac{8}{3}\sum_{i=1}^{N}f^4(i)}>0.\end{gathered} \tag{9}$$

The following theorem states the problem of choosing the threshold constant C in the decision rule:

Theorem 1. If for a certain $\delta > 0$

$$\lim_{N\to\infty}\left\{\sum_{i=1}^{N}f^{4+2\delta}(i)\Bigg/\left[\sum_{i=1}^{N}f^4(i)\right]^{1+\delta/2}\right\}=0, \tag{10}$$

then as $N \to \infty$ under hypothesis H_0, S converges across the distribution to N(0, 1), i.e., to a normal random variable with zero mean value and unit variance.

Without presenting the proof of this theorem, we note only that it reduces to proof of the conditions of the Lyapunov theorem [1].

We can now state the final decision rule, at least for the asymptotic case $N \to \infty$.

Let C_α be determined from the condition

$$\int_{-\infty}^{C_\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \alpha, \tag{11}$$

where α is a given significance level. Then:

- if $S < C_\alpha$, adopt hypothesis H_0;
- if $S \geq C_\alpha$, reject H_0 at significance level α.

We note further that under alternative H_1

$$D\{S_2 | H_1\} = \frac{8}{3}\sum_{i=1}^{N} f^4(i) + 8\sum_{i=1}^{N} f_1^2(i)\,(f(i) - f_1(i))^2 \tag{12}$$

and the following theorem applies:

Theorem 2. If for a certain $\delta > 0$

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N}\left[f^{4+2\delta}(i) + (f_1(i)\cdot f(i))^{2+\delta} + (2f_1(i)\cdot f(i) - f_1^2)^{2+\delta}\right]}{\left[\frac{1}{3}\sum_{i=1}^{N} f_1^4(i) + \sum_{i=1}^{N}(f(i) - f_1(i))^2\right]^{1+\delta/2}} = 0, \tag{13}$$

then under alternative H_1 the statistic

$$\frac{\sum_{i=1}^{N}\left[\frac{x_i^4}{3} - 2x_i^2 f(i)\right] - \sum_{i=1}^{N}(f_1(i) - f(i))^2}{\sqrt{\frac{8}{3}\sum_{i=1}^{N} f^4(i) + 8\sum_{i=1}^{N} f^2(i)\,(f_1(i) - f(i))^2}} \tag{14}$$

converges across the distribution to N(0, 1).

Proof of this theorem also reduces to verification of the Lyapunov—theorem conditions. At least in the asymptotic case $N \to \infty$, it enables to write the form of the power function of the proposed criterion.

TESTING THE HYPOTHESIS THAT THE VARIANCE IS PROPORTIONAL TO A GIVEN FUNCTION

Let us now consider testing of a hypothesis of the form

$$\begin{aligned} &H_0:\ D\{x_i\} = C\cdot f(i),\ i = \overline{1, N}, \\ &H_1:\ D\{x_i\} \neq C\cdot f(i), \end{aligned} \tag{15}$$

where f(i) is a certain function and C is an unknown arbitrary constant. In other words, test the hypothesis that the dependence of $D\{x_i\}$ on i is proportional to the function f(i).

We use the same idea as above in constructing a statistic for verification of this hypothesis: find a statistic S whose mean value is zero under hypothesis H_0 and always greater than zero under the alternative H_1.

We construct this statistic in several steps.

1. Let hypothesis H_0 be true. Since the coefficient C is unknown, it must be estimated. Since $M\{x_i^2\} = C\cdot f(i)$, we use least squares to obtain an estimate $\hat{C}$ of C from the condition

$$R=\sum_{i=1}^{N}(x_i^2-\hat{C}f(i))^2 \Rightarrow \min_{\hat{C}}. \tag{16}$$

Equating the derivative of R with respect to $\hat{C}$ to zero, we obtain

$$\hat{C}=\sum_{i=1}^{N}x_i^2\Big/\sum^{N}f^2(i). \tag{17}$$

2. We take a statistic of the form

$$S_1=\sum_{i=1}^{N}(x_i^2-\hat{C}f(i))^2, \tag{18}$$

which will be corrected later, as a basis for construction of the statistic that we need. For brevity we denote

$$F_\kappa(N)=\sum_{i=1}^{N}f^\kappa(i). \tag{19}$$

Then statistic S_1 can be rewritten in the form

$$S_1=\sum_{i=1}^{N}x_i^4-\frac{1}{F_2(N)}\left(\sum_{i=1}^{N}x_i^2 f(i)\right)^2. \tag{20}$$

Calculating the mean value of this statistic under alternative H_1, when $D\{x_i\} = C\cdot f(i)$, we find that

$$\begin{aligned} M\{S_1\mid H_1\}=C^2\frac{\sum_{i=1}^{N}f_1^2(i)\cdot\sum_{i=1}^{N}f^2(i)-(\sum_{i=1}^{N}f_1(i)f(i))^2}{F_2(N)}+\\ +2C^2\frac{\sum_{i=1}^{N}f_1^2(i)\cdot\sum_{i=1}^{N}f^2(i)-\sum_{i=1}^{N}f_1^2(i)f^2(i)}{F_2(N)}. \end{aligned} \tag{21}$$

The first term satisfies the necessary inequality, since according to the Schwartz inequality [2]

$$\sum_{i=1}^{N}f_1^2(i)\cdot\sum_{i=1}^{N}f^2(i)-(\sum_{i=1}^{N}f_1(i)f(i))^2\geq 0$$

and equality to zero can hold only when $f_i(i)/f(i) = \text{const}$. The second term, on the other hand, is redundant and must be removed.

3. To cancel the second term we note that $M\{x_i^4|H_1\} = 3C^2f_1^2(i)$ and that the mean value of the expression

$$\frac{2}{3F_2(N)}\left[\sum_{i=1}^{N}x_i^4\cdot\sum_{i=1}^{N}f^2(i)-\sum_{i=1}^{N}x_i^4f^2(i)\right]$$

is nothing other than the second term in (21). Subtracting this term from S_1, we obtain the new statistic

$$S_2=\frac{1}{3}\sum_{i=1}^{N}\left(1+2\frac{f^2(i)}{F_2(N)}\right)x_i^4-\frac{1}{F_2(N)}\left(\sum_{i=1}^{N}x_i^2f(i)\right)^2, \tag{22}$$

whose mean value is equal to the first term in (21) and therefore satisfies the necessary condition.

4. For subsequent correction of this statistic we find its variance under hypothesis H_0. Omitting the unwieldy paperwork, we present only the final result:

$$D\{S_2\mid H_0\}=C^4\left[\frac{2}{3}\left(4F_4(N)-15\frac{F_6(N)}{F_2(N)}-\frac{F_8(N)}{F_2^2(N)}+12\frac{F_4^2(N)}{F_2^2(N)}\right)\right]. \tag{23}$$

We now apply the final correction to the statistic. We should like to obtain not only $M\{S|H_0\} = 0$ under this hypothesis, but also $D\{S|H_0\} = 1$. To do this we must divide S_2 by $\sqrt{D\{S_2|H_0\}}$, but then the unknown C^2 appears in the numerator. Therefore C^2 must be replaced by an estimate. Since $M\{x_i^4\} = 3C^2f^2(i)$ under hypothesis H_0, we find the estimate $\hat{C}^2$ of C^2 by least squares from the condition

$$\sum_{i=1}^{N} (x_i^4 - 3\hat{C}^2 f^2(i))^2 \Rightarrow \min_{\hat{C}}.$$

Hence

$$\hat{C}^2 = \sum_{i=1}^{N} x_i^4 f^2(i) / 3F_4(N).$$

Replacing C^2 in (23) with its estimate and dividing S_2 by $\sqrt{D\{S_2|H_0\}}$, we obtain the final expression for the unknown statistic in the form

$$S = K \cdot \frac{\frac{1}{3}\sum_{i=1}^{N}\left(1 + 2\frac{f^2(i)}{F_2(N)}\right)x_i^4 - \frac{1}{F_2(N)}\left(\sum_{i=1}^{N} x_i^2 f(i)\right)^2}{\sum_{i=1}^{N} x_i^4 f^2(i)}; \tag{24}$$

$$K = 3F_4(N) \Big/ \sqrt{\frac{2}{3}\left(4F_4(N) - 15\frac{F_6(N)}{F_2(N)} - \frac{F_8(N)}{F_2^2(N)} + 12\frac{F_4^2(N)}{F_2^2(N)}\right)}. \tag{25}$$

We now present without proof the principal results of analysis of this estimate.

Theorem 3. If as $N \to \infty$, $F_4(N) \uparrow +\infty$ and

$$\sum_{i=1}^{N} \frac{f^8(N)}{F_4^2(N)} < +\infty, \tag{26}$$

then $\hat{C}^2$ converges almost certainly to C^2 as $N \to \infty$.

The proof of this theorem is based on the Gaek—Regnie inequality [3] and Kronecker's lemma [3] in much the same way as Kolmogorov's theorem is proven in [3].

Theorem 4. If there exists a $\delta > 0$ such that

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} f^{4+2\delta}(i) + 2\left(\sum_{i=1}^{N} f^{6+2\delta}(i)\right) \Big/ F_2(N)}{\left(F_4(N) + 4\frac{F_6(N)}{F_2(N)} + 4\frac{F_8(N)}{F_2^2(N)}\right)^{1+\delta/2}} = 0, \tag{27}$$

then as $N \to \infty$ the random variables

$$\xi_N = \frac{\sum_{i=1}^{N}\left(1 + 2\frac{f^2(i)}{F_2(N)}\right)\left(\frac{x_i^4}{3} - C^2 f^2(i)\right)}{C^2\sqrt{\frac{32}{2}\left(F_4(N) + 4\frac{F_6(N)}{F_2(N)} + 4\frac{F_8(N)}{F_2^2(N)}\right)}},$$

$$\eta_N = \left[\sum_{i=1}^{N} (x_i^2 - Cf(i))f(i)\right] / C\sqrt{2F_4(N)} \tag{28}$$

converge across the distribution to a bivariate normal random variable with zero mean values, unit variances, and a correlation coefficient

$$\lim_{N\to\infty}\left(4F_4(N) - 2\frac{F_6(N)}{F_2(N)}\right) \Big/ \sqrt{\frac{4}{3}F_4(N)\left(F_4(N) + 4\frac{F_6(N)}{F_2(N)} + 4\frac{F_8(N)}{F_2^2(N)}\right)}. \tag{29}$$

The proof of this theorem reduces to testing the conditions of the Lyapunov theorem for the random variable $\alpha\xi_N + \beta\eta_N$ with arbitrary α and β. Since α and β are arbitrary, it follows that ξ_N and η_N are jointly normal [4].

These theorems form a basis for proof of the next theorem:

Theorem 5. If the conditions of theorems 3 and 4 and hypothesis H_0 are satisfied, the statistic S converges across the distribution as $N \to \infty$ to N(0, 1), i.e., to a normal random variable with zero mean value and unit variance.

This enables us finally to formulate a criterion for testing of the hypothesis. If it is found that $S < C_\alpha$ (11), adopt hypothesis H_0, but reject it if $S \geq C_\alpha$.

TESTING THE HYPOTHESIS OF STATIONARY WHITE GAUSSIAN NOISE

Consider testing of a hypothesis of the form

$$H_0\colon\ D\{x_i\}=C \quad \forall i=\overline{1,N},$$
$$H_1\colon\ D\{x_i\}\neq C,$$

where C is an unknown constant. This is the hypothesis that the variance of the white Gaussian noise does not vary with time, i.e., the hypothesis that the noise is stationary.

This hypothesis is a particular case of the foregoing hypothesis, from which it can be derived by putting f(i) = 1. Since $F_\kappa(N) = N$ for all κ in that case, the statistic for testing this hypothesis assumes the form

$$S=\frac{\frac{1}{3}\left(1+\frac{2}{N}\right)\sum_{i=2}^{N}x_i^4-\frac{1}{N}\left(\sum_{i=1}^{N}x_i^2\right)^2}{\sum_{i=1}^{N}x_i^4}\,\frac{3N}{\sqrt{\frac{2}{3}\left(4N-3-\frac{1}{N}\right)}}. \tag{30}$$

It is easily verified that all conditions of theorems 3 and 4 are satisfied. Therefore if $S < C_\alpha$, adopt the stationary-noise hypothesis, and if $S > C_\alpha$, reject it.

REFERENCES

1. V. V. Gnedenko, Textbook of Probability Theory (in Russian), Nauka, Moscow (1988).
2. M. Loeve, Probability Theory, Van Nostrand-Reinhold (1963).
3. C. R. Rao, Linear Statistical Inference and Its Applications, Wiley (1965).
4. T. Anderson, Statistical Analysis of Time Series, Wiley (1971).