A NOTE ON THE TEST OF SERIAL CORRELATION COEFFICIENTS

By Masami Ogawara

Meteorological Research Institute, Tokyo

1. Summary. In this note the author points out that in the case of stationary Gaussian Markov process, i.e., autoregressive stochastic process, we can test the serial correlation coefficients by a method based on normal regression theory. Particularly, in the case of simple Markov process, we can find the confidence limits for its autocorrelation coefficient.

In this method, so far as random variables are concerned, not all the information in the original data is used, with a consequent reduction of degrees of freedom. However, the other part of information is introduced as parameters in the distribution functions of random variables and in the statistic useful for tests.

2. Introduction. For the test of the serial correlation coefficient, a method based on its distribution may be orthodox. Up to the present, however, many investigations along this line, e.g. R. L. Anderson [1], M. H. Quenouille [2], P. A. P. Moran [3], T. W. Anderson [4] and others seem to be confined in at least one of the following restrictions:

(1) circular definition,

(2) significance test, i.e., testing the uncorrelatedness of the process,

(3) approximate distribution.

In this paper, we do not use the distribution of a serial correlation coefficient itself, but normal regression theory, and will give the general testing method for an autoregressive stochastic process.

3. Fundamental theorems. The following theorems are fundamental in our method.

Theorem 1. *Let $x_t(t = \cdots, -1, 0, 1, 2, \cdots)$ be a simple Markov process. If the values of $x_{2k-1}(k = 1, 2, \cdots, n+1)$ are fixed, the random variables x_{2k} $(k = 1, 2, \cdots, n)$ are mutually independent.*[1]

This theorem is easily proved from the following facts:

(1) When the value of x_0 is given, $x_1, \cdots, x_n$ are independent of $x_{-1}, x_{-2}, \cdots$.

(2) When x_0 is given, the stochastic sequence $x_1, x_2, \cdots$, is also a simple Markov process for the inversely directed time scale.

Similarly, the following general theorem holds:

Theorem 2. *Let x_t $(t = \cdots, -1, 0, 1, 2, \cdots)$ be a Markov process of order h. Then, if the values of $x_{k(h+1)-h}, \cdots, x_{k(h+1)-1}, x_{k(h+1)+1}, \cdots, x_{k(h+1)+h}$ $(k = 1, 2, \cdots, n)$ are given, the random variables $x_{k(h+1)}$ $(k = 1, 2, \cdots, n)$ are mutually independent.*

[1] This fact has been used by U. V. Linnik (without proof) in his proof of the central limit theorem for simple Markov process. *Izvestiya Akad. Nauk. USSR., Ser. Mat.*, Vol. 13 (1949).

THEOREM 3.[2] *Let x_t $(t = \cdots, -1, 0, 1, \cdots)$ be a stationary Gaussian process. A necessary and sufficient condition that x_t should be a non-singular Markov process of order h is that its autocorrelation coefficients ρ_τ satisfy the finite difference equation*

$$\rho_\tau + a_1\rho_{\tau-1} + \cdots + a_h\rho_{\tau-h} = 0, \qquad \tau = 1, 2, \cdots; a_h \neq 0, \tag{1}$$

where the a's are such that every root of the equation

$$z^h + a_1 z^{h-1} + \cdots + a_{h-1}z + a_h = 0$$

lies within the unit circle.

4. The case of a stationary Gaussian simple Markov process. Let m, σ^2 and $\rho_\tau(\equiv\rho^\tau)$ be the mean, variance and autocorrelation coefficient, respectively, of a stationary Gaussian simple Markov process x_t with discrete parameter t. According to Theorem 1, when the values of $x_{2k-1}(k = 1, 2, \cdots, n + 1)$ are fixed, $x_{2k}(k = 1, 2, \cdots, n)$ are mutually independent and, in this case, their conditional probability densities are given by

$$f(x_{2k} \mid x_{2k-1}, x_{2k+1}) = \frac{1}{\sqrt{2\pi}\,\sigma_0}\exp\left[-\frac{1}{2\sigma_0^2}\{x_{2k} - (a + bx_k')\}^2\right] \qquad (k = 1, 2, \cdots, n),$$

where

$$\begin{aligned} a &= m(1-\rho)^2/(1+\rho^2), \\ b &= 2\rho/(1+\rho^2), \\ \sigma_0^2 &= \sigma^2(1-\rho^2)/(1+\rho^2), \\ x_k' &= (x_{2k-1} + x_{2k+1})/2. \end{aligned} \tag{2}$$

Considering x_k' as the fixed variates and applying normal regression theory [6], the maximum likelihood estimates of the parameter a, b, and σ_0^2 are given by

$$\begin{aligned} \hat{a} &= \bar{x}_2 - \hat{b}\bar{x}', \\ \hat{b} &= \sum_{k=1}^{n}(x_k' - \bar{x}')(x_{2k} - \bar{x}_2) \Big/ \sum_{k=1}^{n}(x_k' - \bar{x}')^2, \\ \hat{\sigma}_0^2 &= \sum_{k=1}^{n}(x_{2k} - \hat{a} - \hat{b}x_k')^2/n, \end{aligned} \tag{3}$$

where

$$\bar{x}_2 = \sum_1^n x_{2k}/n,\ \bar{x}' = \sum_1^n x_k'/n = (\bar{x}_1 + \bar{x}_3)/2$$

[2] M. Ogawara [5].

with

$$\bar{x}_1 = \sum_1^n x_{2k-1}/n, \ \bar{x}_3 = \sum_1^n x_{2k+1}/n.$$

We can rewrite $\hat{b}$ as follows:

(4) $$\hat{b} = 2r_1/(1 + r_2),$$

where

(5)
$$r_1 = \frac{\frac{1}{2}\left\{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)(x_{2k} - \bar{x}_2) + \frac{1}{n}\sum_1^n (x_{2k} - \bar{x}_2)(x_{2k+1} - \bar{x}_3)\right\}}{\frac{1}{2}\left\{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)^2 + \frac{1}{n}\sum_1^n (x_{2k+1} - \bar{x}_3)^2\right\}},$$

$$r_2 = \frac{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)(x_{2k+1} - \bar{x}_3)}{\frac{1}{2}\left\{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)^2 + \frac{1}{n}\sum_1^n (x_{2k+1} - \bar{x}_3)^2\right\}}.$$

Because

$$\frac{\partial(a, b, \sigma_0^2)}{\partial(m, \sigma^2, \rho)} = \frac{2(1 - \rho)^2(1 - \rho^2)^2}{(1 + \rho^2)^4} > 0 \qquad (\text{for } |\rho| \neq 1),$$

the maximum likelihood estimates of m, σ^2 and ρ are given by

(6)
$$\hat{m} = \hat{a}/(1 - \hat{b}),$$
$$\hat{\sigma}^2 = \hat{\sigma}_0^2/\sqrt{1 - \hat{b}^2},$$
$$\hat{\rho} = (1 - \sqrt{1 - \hat{b}^2})/\hat{b}.$$

Since, as the function of random variables x_{2k},

(7) $$F = \frac{(\hat{b} - b)^2 \sum_1^n (x_k' - \bar{x}')^2 \cdot (n - 2)}{\sum_1^n (x_{2k} - \hat{a} - \hat{b}x_k')^2}$$

has the F-distribution with 1 and $n - 2$ degrees of freedom, we can test the hypotheses $\rho = \rho_0$ or $b = b_0 = 2\rho_0/(1 + \rho_0^2)$. As the function $\rho = (1 - \sqrt{1 - b^2})/b$ is monotone increasing, we can also find confidence limits for ρ from those for b.

5. The case of a stationary Gaussian Markov process of order h. Let, as before, m, σ^2 and ρ_τ be the mean, variance and autocorrelation coefficient of our process x_t respectively. From Theorem 2, the random variables $x_{k(h+1)}(k = 1, 2, \cdots, n)$ are independent of each other, under the condition that the variables $x_{k(h+1)-p}$, $x_{k(h+1)+p}$ $(p = 1, 2, \cdots, h;\ k = 1, 2, \cdots, n)$ are fixed, and, in the present case, their conditional probability densities are given by

(8)
$$f(x_{k(h+1)} \mid x_{k(h+1)-p}, x_{k(h+1)+p};\ p = 1, 2, \cdots, h)$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma_0^2} \exp\left[-\frac{1}{2\sigma_0^2}\left\{x_{k(h+1)-p} - \sum_{p=0}^{h} b_p x_{pk}'\right\}^2\right] \qquad (k = 1, 2, \cdots, n),$$

where $x'_{pk} = (x_{k(h+1)-p} + x_{k(h+1)+p})/2$ $(p = 1, 2, \cdots, h)$, $x'_{0k} = 1$, and where

$$b_0 = m\left(1 - 2\sum_{p=1}^{h} c_p\right), \quad b_p = 2c_p \ (p = 1, 2, \cdots, h),$$

$$\begin{pmatrix} c_h \\ c_{h-1} \\ \vdots \\ c_1 \\ c_1 \\ \vdots \\ c_h \end{pmatrix} = \begin{pmatrix} 1 & \cdots & \rho_{h-1} & \rho_{h+1} & \cdots & \rho_{2h} \\ \vdots & & \vdots & \vdots & & \vdots \\ \rho_{h-1} & \cdots & 1 & \rho_2 & \cdots & \rho_{h+1} \\ \rho_{h+1} & \cdots & \rho_2 & 1 & \cdots & \rho_{h-1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \rho_{2h} & \cdots & \rho_{h+1} & \rho_{h-1} & \cdots & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \rho_h \\ \rho_{h-1} \\ \vdots \\ \rho_1 \\ \rho_1 \\ \vdots \\ \rho_h \end{pmatrix} \tag{9}$$

and

$$\sigma_0^2 = \frac{1 + a_1\rho_1 + \cdots + a_h\rho_h}{1 + a_1^2 + \cdots + a_h^2}\,\sigma^2, \tag{10}$$

where the a's are the coefficients of equation (1).

Considering the relations (1) and (9), the hypotheses concerning $\rho_1, \cdots, \rho_h$[3] is equivalent to the hypotheses concerning $c_1, \cdots, c_h$ or $b_1, \cdots, b_h$. Thus normal regression theory is applicable.

Moreover, we can estimate the order of the Markov process as follows. The above stated theory holds whenever the essential order h_0 of the process is not greater than h. Hence, we may select as its order such a value h_0 that the hypotheses $b_{h_0} = b_{h_0+1} = \cdots = b_h = 0$ is rejected but the hypotheses $b_{h_0+1} = b_{h_0+2} = \cdots = b_h = 0$ is not rejected.

REFERENCES

[1] R. L. Anderson, "Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1–13.

[2] M. H. Quenouille, "Some results in the testing of serial correlation coefficients," *Biometrika*, Vol. 35 (1948), pp. 261–267.

[3] P. A. P. Moran, "Some theorems on time series, II. The significance of the serial correlation coefficient," *Biometrika*, Vol. 35 (1948), pp. 255–260.

[4] T. W. Anderson, "On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, Vol. 31 (1948), pp. 88–116.

[5] M. Ogawara, "On the normal stationary Markov process of higher order," *Bull. of Math. Stat.* (in Japanese), Vol. 2 (1948), pp. 42–46.

[6] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1943.

[3] Owing to (1), this is also equivalent to the hypotheses concerning $a_1, \cdots, a_h$.