Aplikace matematiky

Jan Hurt

On a simple estimate of correlations of stationary random sequences

Aplikace matematiky, Vol. 18 (1973), No. 3, 176–187

Persistent URL: http://dml.cz/dmlcz/103468

ON A SIMPLE ESTIMATE OF CORRELATIONS OF STATIONARY RANDOM SEQUENCES

JAN HURT

(Received March 29, 1972)

This paper deals with a simple estimator for the correlation function of a stationary Gaussian random sequence. In Section 1 the assumptions are formulated and a statistic based on signs of original values is proposed. The basic properties of that statistic such as its expectation and variance are given in Section 2. On the basis of these properties, the proposed statistic can be used to estimate correlations of stationary Gaussian random sequences. In Section 3 the asymptotic normality of the discussed statistic is proved. The last section of the present paper contains some numerical results for stationary Gaussian autoregressive series.

1. PRELIMINARIES

Suppose that $\{X_t\}_{t\in T}$ is a (weakly) stationary Gaussian discrete random process where T is the set of integers. Assume $EX_t = 0$, $t \in T$. Let $\{\varrho_j\}_{j=0}^{\infty}$ be the correlation function of $\{X_t\}_{t\in T}$, i.e., $\varrho_j = EX_tX_{t+j}/\sigma^2$ where $\sigma^2 = EX_0^2$. Define the sequence $\{Z_t\}_{t\in T}$ by

(1) $$Z_t = \operatorname{sign} X_t\,, \quad t \in T$$

and put

(2) $$T_{tj} = Z_tZ_{t+j}\,, \quad t \in T$$

for j natural. The quantities T_{tj} will be used to estimate the correlation function of the process $\{X_t\}_{t\in T}$.

2. BASIC PROPERTIES

From the normality of the marginal distributions of X_t, $t \in T$, it follows that $P(Z_t = 0) = 0$, $t \in T$. Hence the event $(Z_t = 0)$ may be neglected in our considerations. Now we derive the distribution of the variable T_{tj}. The quantity T_{tj} assumes only the values $+1$, -1 with nonzero probability. Thus

$$P(T_{tj} = 1) = P(Z_t = 1, Z_{t+j} = 1) + P(Z_t = -1, Z_{t+j} = -1).$$

Random variables $X_t, \ldots, X_{t+N-1}$ $(N > j)$ have the simultaneous Gaussian distribution with vanishing means and the covariance matrix $\mathbf{G} = (\sigma_{ik})_{i,k=1}^{N}$ where $\sigma_{ik} = \varrho_{|i-k|}\,\sigma^2$. Therefore the joint distribution of X_t, X_{t+j} is Gaussian with the covariance matrix

$$\mathbf{G}_j = \sigma^2 \begin{bmatrix} \varrho_j & 1 \\ 1 & \varrho_j \end{bmatrix}.$$

Integrating the joint density of X_t, X_{t+j} over $(0, \infty) \times (0, \infty)$, we can express the probability $P(Z_t = 1, Z_{t+j} = 1)$ as follows:

(3) $$P(Z_t = 1, Z_{t+j} = 1) = (2\pi)^{-1} (\pi - \arccos \varrho_j) = (2\pi)^{-1} \arccos(-\varrho_j).$$

Similarly, we can obtain

(4) $$P(Z_t = -1, Z_{t+j} = -1) = (2\pi)^{-1} \arccos(-\varrho_j)$$

so that

(5) $$P(T_{tj} = 1) = \pi^{-1} \arccos(-\varrho_j).$$

Therefore

(6) $$P(T_{tj} = -1) = 1 - P(T_{tj} = 1) = \pi^{-1} \arccos \varrho_j.$$

Further,

(7) $$ET_{tj} = \pi^{-1}\, 2 \arccos(-\varrho_j) - 1 = \pi^{-1}\, 2 \arcsin \varrho_j$$

and

(8) $$\operatorname{var} T_{tj} = 4\left[\pi^{-1} \arccos \varrho_j - (\pi^{-1} \arccos \varrho_j)^2\right].$$

If $\hat{\mu}_j$ is an estimator for ET_{tj}, we can take the quantity

(9) $$\hat{\varrho}_j = \sin\left(\frac{\pi}{2}\hat{\mu}_j\right)$$

as an estimator for ϱ_j. We shall see below that

$$\bar{T}_i = \frac{1}{N - j} \sum_{t=1}^{N-j} T_{tj}$$

is an estimator for ET_{0j} which has some appropriate properties.

First of all we see that $\bar{T}_j$ is the unbiased estimator for ET_{tj}. The expression for the variance of $\bar{T}_j$ is rather complicated. To derive it we use the following lemmas.

Lemma 1. *Any Gaussian stationary discrete process is strictly stationary.*

Proof. The assertion follows from the fact that the joint Gaussian distribution is fully determined by the second order moments.

Lemma 2. *The random process* $\{Z_t\}_{t\in T}$ *is strictly stationary.*

Proof. Let $y_{t_1}, \ldots, y_{t_n}$ be an arbitrary sequence of numbers $+1, -1$. Let h be integer and $t_1, \ldots, t_m \in T$. Then

$$P(Z_{t_1} = y_{t_1}, \ldots, Z_{t_m} = y_{t_m}) = P(y_{t_1}X_{t_1} > 0, \ldots, y_{t_m}X_{t_m} > 0) =$$
$$= P(y_{t_1}X_{t_1+h} > 0, \ldots, y_{t_m}X_{t_m+h} > 0) = P(Z_{t_1+h} = y_{t_1}, \ldots, Z_{t_m+h} = y_{t_m}) .$$

Lemma 3. *Let j be a fixed natural number. Then the random process* $\{T_{tj}\}_{t\in T}$ *is strictly stationary.*

Proof. The proof is similar to that of Lemma 2.

In view of Lemma 3 the quantity $\operatorname{cov}(T_{tj}, T_{sj})$ depends only on the difference $t - s$. Denote

(10) $$R_j(t - s) = \operatorname{cov}(T_{tj}, T_{sj})$$

for fixed natural j. After some computation it may be shown that

(11) $$\operatorname{var} \sum_{t=1}^{N-j} T_{tj} = (N - j)\, R_j(0) + \sum_{k=1}^{N-j-1} (N - j - k)\, R_j(k) .$$

Let us evaluate the covariances $R_j(k)$. Because of the stationarity,

(12) $$R_j(k) = ET_{0j}T_{kj} - (ET_{0j})^2 .$$

Obviously

(13) $$\begin{aligned} ET_{0j}T_{kj} &= P(T_{0j}T_{kj} = 1) - P(T_{0j}T_{kj} = -1) \\ &= 2P(T_{0j}T_{kj} = 1) - 1 . \end{aligned}$$

Denote by A_t the event $(Z_t = 1) = (X_t > 0)$ and by $\bar{A}_t$ its complement.

Then

(14) $$\begin{aligned} P(T_{0j}T_{kj} = 1) &= P(Z_0 Z_j Z_k Z_{k+j} = 1) = \\ &= P(A_0 A_j A_k A_{k+j}) + P(A_0 A_j \bar{A}_k \bar{A}_{k+j}) + \\ &+ P(A_0 \bar{A}_j A_k \bar{A}_{k+j}) + P(A_0 \bar{A}_j \bar{A}_k A_{k+j}) + \\ &+ P(\bar{A}_0 A_j A_k \bar{A}_{k+j}) + P(\bar{A}_0 A_j \bar{A}_k A_{k+j}) + \\ &+ P(\bar{A}_0 \bar{A}_j A_k A_{k+j}) + P(\bar{A}_0 \bar{A}_j \bar{A}_k \bar{A}_{k+j}) . \end{aligned}$$

Using the well-known formula

$$P(AB\bar{C}D) = P(AB\bar{C}) - P(AB\bar{C}D) = P(AB) - P(ABC) - P(ABD) + P(ABCD) \tag{15}$$

which takes place for any random events A, B, C, D, the formula (14) may be simplified and we obtain

$$\begin{aligned} P(T_{0j}T_{kj} = 1) = {} & 8P(A^+) + 2P(A_0A_j) + 2P(A_0A_k) + \\ & + 2P(A_0A_{k+j}) + 2P(A_jA_k) + \\ & + 2P(A_jA_{k+j}) + 2P(A_kA_{k+j}) - \\ & - 4P(A_0A_jA_k) - 4P(A_0A_jA_{k+j}) - \\ & - 4P(A_0A_kA_{k+j}) - 4P(A_jA_kA_{k+j}) - 1 \end{aligned} \tag{16}$$

where $A^+ = A_0A_jA_kA_{k+j}$.

An explicit formula for calculating $P(A^+)$ is not known in general case. Plackett has derived a reduction formula for normal multivariate integrals (see [6]). His procedure may be used in order to calculate $P(A^+)$ numerically. In the case that the event E is an intersection of at most three of the events A_0, A_j, A_k, A_{k+j}, the same author has derived explicit formulas for $P(E)$ (see again [6]).

The correlation matrix of (X_0, X_j, X_k, X_{k+j}) is

$$\begin{bmatrix} 1 & \varrho_j & \varrho_k & \varrho_{k+j} \\ & 1 & \varrho_{k-j} & \varrho_k \\ & & 1 & \varrho_j \\ & & & 1 \end{bmatrix}. \tag{17}$$

From (3) it follows that

$$P(A_tA_{t+s}) = (2\pi)^{-1} \arccos(-\varrho_s) \tag{18}$$

and Plackett's formula for the three-dimensional normal integral on the range $(0, \infty) \times (0, \infty) \times (0, \infty)$ gives

$$\begin{aligned} & P(A_tA_{t+s}A_{t+s+u}) = \\ & = (4\pi)^{-1} [\arccos(-\varrho_s) + \arccos(-\varrho_{s+u}) + \arccos(-\varrho_u) - \pi]. \end{aligned} \tag{19}$$

Using (13), (18), and (19) we get

$$\begin{aligned} ET_{0j}T_{kj} = {} & 16P(A^+) - 2\pi^{-1} [2\arccos(-\varrho_j) + 2\arccos(-\varrho_k) + \\ & + \arccos(-\varrho_{k+j}) + \arccos(-\varrho_{k-j})] + 5. \end{aligned} \tag{20}$$

Finally,

$$\begin{aligned} R_j(k) = {} & 16P(A^+) - 2\pi^{-1} [2\arccos(-\varrho_k) + \arccos(-\varrho_{k+j}) + \\ & + \arccos(-\varrho_{k-j})] - 4\pi^{-2} \arccos^2(-\varrho_j) + 4. \end{aligned} \tag{21}$$

Sometimes, it will be more convenient to use the expression

$$R_j(k) = 4[4P(A^+) - \pi^{-2} \arccos^2(-\varrho_j) - \pi^{-1} \arcsin \varrho_k - (2\pi)^{-1} \arcsin \varrho_{k+j} - (2\pi)^{-1} \arcsin \varrho_{k-j}] \tag{22}$$

which follows from (21) in view of $\arccos(-x) = \frac{1}{2}\pi + \arcsin x$. Note that the formulas hold for any integer k.

3. ASYMPTOTIC PROPERTIES

In this section we shall prove the asymptotic normality of $\overline{T}_j$ for a wide class of random sequences and the law of large numbers for $\{T_{tj}\}_{t\in T}$. Our method is based on some results of Ibragimov and Linnik ([3]).

We say that a strictly stationary random sequence $\{X_t\}_{t\in T}$ satisfies the *strong mixing* condition if

$$\alpha_X(\tau) = \sup_{A \in \mathfrak{M}^0_{-\infty}(X),\, B \in \mathfrak{M}^{\infty}_{\tau}(X)} |P(AB) - P(A)\,P(B)| \to 0 \tag{23}$$

for $\tau \to \infty$, $\tau > 0$, where $\mathfrak{M}^b_a(X)$ is the minimal σ-field generated by the class $\{A\}$ of sets in Ω,

$$A = \{\omega : (X_{t_1}(\omega), \ldots, X_{t_s}(\omega)) \in B_s\}\,, \quad B_s \in \mathfrak{B}_s\,, \quad a \leqq t_1 < \ldots < t_s \leqq b,\ t_1, \ldots, t_s \in T\,.$$

($\mathfrak{B}_s$ denotes the σ-field of Borel sets of R_s.)

Lemma 4. *If $\{X_t\}_{t\in T}$ is a random sequence satisfying the strong mixing condition then $\{Z_t\}_{t\in T}$ satisfies the strong mixing condition, too.*

Proof. Suppose $a \leqq t_1 < \ldots < t_s \leqq b$, $t_1, \ldots, t_s \in T$. The mapping $(X_{t_1}, \ldots, X_{t_s})$: $\Omega \to R_s$ is $\mathfrak{M}^b_a(X) - \mathfrak{B}_s$-measurable. The function sign: $R \to R$ is measurable so that the composite mapping sign X_t: $\Omega \to R$ is $\mathfrak{M}^b_a(X) - \mathfrak{B}$-measurable, $a \leqq t \leqq b$. In view of Theorem II.5.3.b ([5]) the mapping

(sign $X_{t_1}, \ldots,$ sign X_{t_s}): $\Omega \to R_s$ is $\mathfrak{M}^b_a(X) - \mathfrak{B}_s$-measurable so that $(Z_{t_1}, \ldots, Z_{t_s})^{-1}(B) \in \mathfrak{M}^b_a(X)$ for each $B \in \mathfrak{B}_s$. Because of the minimality of $\mathfrak{M}^b_a(Z)$ it is $\mathfrak{M}^b_a(Z) \subset \mathfrak{M}^b_a(X)$, particularly $\mathfrak{M}^0_{-\infty}(Z) \subset \mathfrak{M}^0_{-\infty}(X)$, $\mathfrak{M}^{\infty}_{\tau}(Z) \subset \mathfrak{M}^{\infty}_{\tau}(X)$. Therefore $\alpha_X(\tau) \geqq \alpha_Z(\tau)$ for all $\tau > 0$.

Lemma 5. *If $\{Z_t\}_{t\in T}$ is a random sequence satisfying the strong mixing condition then $\{T_{tj}\}_{t\in T}$ satisfies the strong mixing condition for any j natural.*

Proof. For $a \leqq t_1 < \ldots < t_s \leqq b$ the mapping $(T_{t_1 j}, \ldots, T_{t_s j})$: $\Omega \to R_s$ is $\mathfrak{M}^b_a(T_j) - \mathfrak{B}_s$-measurable. Both Z_t: $\Omega \to R$ and Z_{t+j}: $\Omega \to R$ are $\mathfrak{M}^{b+j}_a(Z) - \mathfrak{B}$-

measurable so that $Z_t Z_{t+j}$: $\Omega \to R$ is $\mathfrak{M}_a^{b+j}(Z)$ – $\mathfrak{B}$-measurable. Therefore $(T_{t_1 j}, \ldots, T_{t_s j})$: $\Omega \to R_s$ is $\mathfrak{M}_a^{b+j}(Z)$ – $\mathfrak{B}_s$-measurable. This implies that $\mathfrak{M}_a^b(T_j) \subset \mathfrak{M}_a^{b+j}(Z)$, $\alpha_{T_j}(\tau) \leqq \alpha_Z(\tau - j)$ for $\tau > j$ and $0 \leqq \alpha_{T_j}(\tau) \leqq \alpha_Z(\tau - j) \to 0$ for $\tau \to \infty$.

Note that the important inequalities

$$\alpha_X(\tau) \geqq \alpha_Z(\tau) \geqq \alpha_{T_j}(\tau + j) \tag{24}$$

for the strong mixing coefficients follow from the proofs of Lemmas 4,5.

Properties of the strong mixing coefficients play a crucial role in the central limit theorems for dependent random variables. However, the investigation of the properties of the coefficient $\alpha_X(\tau)$ is somewhat difficult, in general. Kolmogorov and Rozanov (see e.g. [7]) have derived an important result concerning the behaviour of $\alpha_X(\tau)$ for Gaussian stationary random sequences. Let $-\infty \leqq a \leqq b \leqq \infty$. Denote by $L_a^b(X)$ the closed linear hull generated by $X_a, \ldots, X_b$. Here the convergence in the mean is considered.
Define

$$\varrho_X(\tau) = \sup_L EUV \tag{25}$$

where $L = \{(U, V)\colon U \in L_{-\infty}^0(X),\ V \in L_\tau^\infty(X),\ EU^2 = EV^2 = 1\}$. Then for Gaussian stationary random sequences $\{X_t\}_{t\in T}$ the inequality

$$\alpha_X(\tau) \leqq \varrho_X(\tau) \leqq 2\pi\alpha_X(\tau) \tag{26}$$

holds. (See [7]).

Theorem 1. *Suppose that $\{X_t\}_{t\in T}$ is a Gaussian stationary random sequence with the zero mean and the spectral density $f(\lambda) > m > 0$ where m is a constant. Suppose that f is k-times differentiable and $|f^{(k)}(\lambda)| \leqq M_k$ $(k > 1)$.*
Then

$$\varrho_X(\tau + 1) \leqq A M_k \tau^{-k} \ln \tau\, m^{-1}\,, \quad (\tau \geqq 2)$$

where A is a constant which does not depend on f. Moreover, if $f^{(k)}(\lambda)$ is a function of bounded variation on $[-\pi, \pi]$ then

$$\varrho_X(\tau + 1) \leqq \tau^{-k} V_k m^{-1} \quad (\tau \geqq 2)\,. \tag{27}$$

Proof. The proof of the above assertion may be obtained by a slight modification of the proof of Theorem 17.7.3 in Ibragimov's and Linnik's book ([3], see also [2]).

Theorem 2. *Any Gaussian stationary autoregressive random sequence satisfies the strong mixing condition.*

Proof. Let us consider a Gaussian stationary autoregressive random sequence $\{X_t\}_{t\in T}$ generated by

$$\sum_{k=0}^{n} a_{n-k} X_{t-k} = Y_t,\ t \in T\,,$$

where $\{Y_t\}_{t\in T}$ is a white noise with unit variance. The spectral density function of such a sequence is

$$f(\lambda) = \frac{1}{2\pi}\left|\sum_{k=0}^{n} a_{n-k}\, e^{ik\lambda}\right|^{-2}, \quad \lambda \in [-\pi, \pi].$$

The stationarity of $\{X_t\}_{t\in T}$ guarantees $\sum a_{n-k}\, e^{ik\lambda} \neq 0$ so that $f(\lambda)$ is continuous on $[-\pi, \pi]$ and attains a minimum $m > 0$ on $[-\pi, \pi]$. After the explicit evaluation of $f''(\lambda)$ it may be shown that $|f''(\lambda)|$ is bounded. Therefore we can apply Theorem 1 to $\{X_t\}_{t\in T}$. Further $\tau^{-1} \ln \tau \to 0$ for $\tau \to \infty$ which proves the theorem.

To prove the asymptotic normality of $N^{-1} \sum_{t=1}^{N} T_{tj}$ we make use of the following formulation of the central limit theorem:

Theorem 3. *Let $\{W_t\}_{t\in T}$ be a strictly stationary random sequence with the mean zero. Suppose that $\{W_t\}_{t\in T}$ satisfies the strong mixing condition with a coefficient $\alpha_W(\tau)$. Suppose $\sum_{\tau=1}^{\infty} \alpha_W(\tau) < \infty$. Let W_t be uniformly bounded with probability one. Then*

$$\sigma_W^2 = EW_0^2 + 2\sum_{t=1}^{\infty} EW_0W_t < \infty\,. \tag{28}$$

If $\sigma_W \neq 0$ then

$$\lim_{N\to\infty} P\left\{\frac{1}{\sigma_W\sqrt{N}} \sum_{t=1}^{N} W_t < z\right\} = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-x^2/2}\, dx\,. \tag{29}$$

Proof. See [3], Theorem 18.5.4.

Theorem 4. *Let $\{X_t\}_{t\in T}$ be a Gaussian stationary autoregressive random sequence. Then*

$$\sigma_{T_j}^2 = R_j(0) + 2\sum_{k=1}^{\infty} R_j(k) < \infty \tag{30}$$

for fixed j natural and the limiting distribution of

$$\sqrt{N}\,(\overline{T}_j - ET_{0j}) \text{ as } N \to \infty \text{ is } N(0, \sigma_{T_j}^2)\,.$$

Proof. Put $T_{tj}^* = T_{tj} - ET_{0j}$ for $t \in T$.
The random sequence $\{T_{tj}\}_{t\in T}$ is strictly stationary with the zero means and satisfies the strong mixing condition with a coefficient $\alpha_{T_j^*}(\tau) = \alpha_{T_j}(\tau)$. Further $|T_{tj}^*| < 1 + ET_{0j}$ with probability one. It follows from Theorem 1 that $\varrho_X(\tau + 1) \leqq$ const. $\tau^{-2} \ln \tau$. From (24) and (26) we obtain $\alpha_{T_j}(\tau + 1 + j) \leqq \varrho_X(\tau + 1)$ so that $\alpha_{T_j^*}(\tau + 1 + j) \leqq$ const. $\tau^{-2} \ln \tau$ for $\tau \geqq 2$. The series $\sum_{\tau=2}^{\infty} \tau^{-2} \ln \tau$ converges, hence

$\sum_{\tau=2}^{\infty} \alpha_{T_j}(\tau + 1 + j) < \infty$. Obviously $\sigma_{T_j}^2 = \sigma_{T_j*}^2$. Therefore, it follows from Theorem 3 that $\sigma_{T_j}^2 < \infty$. If $\sigma_{T_j}^2 > 0$ then the statistic

$$\frac{1}{\sigma_{T_j}\sqrt{N}} \sum_{t=1}^{N} T_{tj}^*$$

has an asymptotically normal distribution $N(0,1)$. It remains to prove that if $\sigma_{T_j}^2 = 0$ then the distribution of $N^{-1/2} \sum T_{tj}^*$ converges to a degenerate one with a saltus at the point zero. It follows from (11) that

$$\text{(31)} \qquad \operatorname{var} \sum_{t=1}^{N} T_{tj} = N[R_j(0) + 2 \sum_{k=1}^{N-1} R_j(k)] - 2 \sum_{k=1}^{N-1} kR_j(k) .$$

Ibragimov and Linnik ([3], p. 388) have shown $|R_j(k)| \leqq 4\alpha_{T_j}(k)$ and thus

$$\text{(32)} \qquad |\sum_{k=1}^{N-1} kR_j(k)| \leqq \sum_{k=1}^{N-1} k\,|R_j(k)| \leqq 4 \sum_{1 \leqq k \leqq \sqrt{N}} k\alpha_{T_j}(k) + 4 \sum_{\sqrt{N} < k \leqq N} k\alpha_{T_j}(k) \leqq$$
$$\leqq 4\sqrt{N} \sum_{1 \leqq k \leqq \sqrt{N}} \alpha_{T_j}(k) + 4N \sum_{\sqrt{N} < k \leqq N} \alpha_{T_j}(k) = o(N) .$$

Consequently $\sigma_{T_j}^2$ and $\operatorname{var} \sum_{t=1}^{N} T_{tj}$ are connected by

$$\text{(33)} \qquad \operatorname{var} \sum_{t=1}^{N} T_{tj} = N\sigma_{T_j}^2(1 + o(1)) .$$

Applying Chebyshev's inequality we obtain

$$P\left\{\left|\frac{1}{\sqrt{N}} \sum_{t=1}^{N} T_{tj}^*\right| < \varepsilon\right\} \geqq 1 - \frac{\sigma_{T_j}^2(1 + o(1))}{\varepsilon^2} \to 1$$

as $N \to \infty$ for any $\varepsilon > 0$.

The following corollary may be obtained in the same manner as the second part of the proof of Theorem 4 (case $\sigma_{T_j} = 0$).

Corollary. *If $\{X_t\}_{t\in T}$ is a Gaussian stationary autoregressive sequence then* $\lim_{N\to\infty} P\{|N^{-1} \sum_{t=1}^{N} T_{tj} - ET_{0j}| < \varepsilon\} = 1$ *for any* $\varepsilon > 0$.

We give a somewhat more general formulation of Theorem 4 (without the assumption of autoregressive model). Its proof is analogous to that of Theorem 4.

Theorem 4'. *Let $\{X_t\}_{t\in T}$ be a Gaussian stationary random sequence with the spectral density $f(\lambda)$. Suppose $f(\lambda) > m > 0$ and $|f''(\lambda)| \leqq M_2$ for $\lambda \in [-\pi,\pi]$. Then (for fixed j natural) the limiting distribution of $\sqrt{N}(\overline{T}_j - ET_{0j})$ as $N \to \infty$ is $N(0, \sigma_{T_j}^2)$.*

4. NUMERICAL RESULTS

We have said that

$$\hat{\varrho}_j = \sin\left(\tfrac{1}{2}\pi \bar{T}_j\right)$$

may serve as an appropriate estimator for ϱ_j. It is clear that $\hat{\varrho}_j$ has the asymptotic distribution

$$\Phi\left[N\,\sigma_{T_j}^{-1}\left(\frac{2}{\pi}\arcsin x - ET_{0j}\right)\right]$$

where Φ is the distribution function of $N(0, 1)$.

The properties of the proposed statistic were investigated numerically for selected stationary autoregressive sequences. Some of these series were simulated on a digital computer. The results are summarized in Tables 1–5. Table 5 contains variances of $\bar{T}_j$ for selected autoregressive series. This table shows a favourable property of var $\bar{T}_1$ in the first order autoregressive series, namely, that var $\bar{T}_1$ does not depend practically on ϱ.

Example 1. 100 series of the type $X_t = 0.6\,X_{t-1} + Y_t$ were generated, each of them of the length 20. Disturbances Y_t are normal $N(0, 1)$. Sample means and sample variances of r_j, $\bar{T}_j$, $\hat{\varrho}_j$ are given in Table 1. This table also contains the theoretical values of ϱ_j to compare ϱ_j with $\hat{\varrho}_j$.

TABLE 1

Sample means and variances of r_j, $\bar{T}_j$, $\hat{\varrho}_j$ for the first order autoregression $X_t = 0{\cdot}6\,X_{t-1} + Y_t$

j	True value of		Mean of			Variance of			Bias of		
	ϱ_j	ET_{0j}	r_j	$\bar{T}_j$	$\hat{\varrho}_j$	r_j	$\bar{T}_j$	$\hat{\varrho}_j$	r_j	$\bar{T}_j$	$\hat{\varrho}_j$
1	·600	·410	·541	·417	·582	·042	·038	·057	−·059	·007	−·018
2	·360	·234	·325	·245	·355	·062	·047	·091	−·035	·009	−.005
3	·216	·138	·210	·158	·230	·070	·065	·128	−·006	·020	·014

Example 2. 50 series of the first order autoregression $X_t = -0.5\,X_{t-1} + Y_t$ were generated, each of them of the length 20. Disturbances Y_t are $N(0, 1)$. Characteristics, the same as in Example 1, are given in Table 2.

Example 3. 20 series of the type $X_t = 0.9\,X_{t-1} + Y_t$ were generated, each of them of the length 100. Correlation characteristics are presented in Table 3.

Example 4. Let us study the behaviour of ϱ_j when the assumption of normality is omitted. The selected series consists of the first 100 terms of Kendall's artificial series 5a ([4]) generated by $X_t = 1.1\,X_{t-1} - 0.5\,X_{t-2} + Y_t$.

TABLE 2

Sample means and variances of r_j, $\bar{T}_j$, $\hat{\varrho}_j$ for the first order autoregression $X_t = -0{\cdot}5\, X_{t-1} + Y_t$

j	True value of		Mean of			Variance of			Bias of		
	ϱ_j	ET_{0j}	r_j	$\bar{T}_j$	$\hat{\varrho}_j$	r_j	$\bar{T}_j$	$\hat{\varrho}_j$	r_j	$\bar{T}_j$	$\hat{\varrho}_j$
1	—·500	—·333	—·489	—·359	—·501	·042	·050	·086	·011	—·026	—·001
2	·250	·161	·247	·191	·272	·051	·059	·112	—·003	·030	·022
3	—·125	—·079	—·122	—·120	—·159	·067	·064	·134	·003	—·041	—·034

TABLE 3

Sample means and variances of r_j, $\bar{T}_j$, $\hat{\varrho}_j$ for the first order autoregression $X_t = 0{\cdot}9\, X_{t-1} + Y_t$

j	True value of		Mean of			Variance of			Bias of		
	ϱ_j	ET_{0j}	r_j	$\bar{T}_j$	$\hat{\varrho}_j$	r_j	$\bar{T}_j$	$\hat{\varrho}_j$	r_j	$\bar{T}_j$	$\hat{\varrho}_j$
1	·900	·713	·870	·706	·882	·002	·010	·006	—·030	—·007	—·018
2	·810	·601	·792	·619	·820	·004	·007	·005	—·018	·018	·010
3	·729	·520	·710	·541	·743	·005	·009	·008	—·019	·021	·014

TABLE 4

Correlation characteristics of Kendall's artificial series $X_t = 1{\cdot}1\, X_{t-1} - 0{\cdot}5\, X_{t-2} + Y_t$

j	ϱ_j	r_j	$\bar{T}_j$	$\hat{\varrho}_j$
1	·733	·787	·596	·805
2	·307	·479	·347	·518
3	—·029	·188	·155	·240
4	—·185	·055	·063	·098
5	—·189	·006	—·074	—·115
6	—·116	·005	—·064	—·100
7	—·033	·004	—·075	—·118
8	·022	·031	·000	·000
9	·041	—·017	—·011	—·017
10	·034	—·010	·044	·070

The disturbances y_t are rectangular with zero means and unit variances in this case. The results are presented in Table 4. It may be seen that the non-normality of Y_t proved not to be significant.

TABLE 5

Variances of $\overline{T}_j$ for selected autoregressive series

A white noise
B first order autoregression $X_t = 0{\cdot}9\, X_{t-1} + Y_t$
C first order autoregression $X_t = 0{\cdot}6\, X_{t-1} + Y_t$
D first order autoregression $X_t = -0{\cdot}5\, X_{t-1} + Y_t$
E second order autoregression $X_t = 1{\cdot}1\, X_{t-1} - 0{\cdot}5\, X_{t-2} + Y_t$

Length of series N	A	B			E
	var $\overline{T}_1$	var $\overline{T}_1$	var $\overline{T}_2$	var $\overline{T}_3$	var $\overline{T}_1$
5	·2500	·1750	·3216	·5009	·1375
10	·1111	·0937	·1608	·2290	·0559
15	·0714	·0652	·1098	·1526	·0350
20	·0526	·0501	·0836	·1149	·0255
25	·0417	·0408	·0675	·0921	·0201
30	·0345	·0344	·0566	·0768	·0166
35	·0294	·0297	·0487	·0659	·0141
40	·0256	·0262	·0427	·0576	·0123
45	·0227	·0234	·0381	·0512	·0109
50	·0204	·0212	·0343	·0461	·0098
55	·0182	·0194	·0313	·0419	·0089
60	·0169	·0178	·0287	·0384	·0082
65	·0156	·0165	·0265	·0355	·0075
70	·0145	·0154	·0247	·0329	·0070
75	·0135	·0144	·0230	·0308	·0065
80	·0127	·0136	·0216	·0288	·0061
85	·0119	·0128	·0204	·0271	·0058
90	·0112	·0122	·0193	·0256	·0055
95	·0106	·0116	·0183	·0243	·0052
100	·0101	·0110	·0174	·0231	·0049

TABLE 5 (continued)

Length of series N	C			D
	var $\overline{T}_1$	var $\overline{T}_2$	var $\overline{T}_3$	var $\overline{T}_1$
5	·2385	·3798	·5680	·2437
10	·1102	·1541	·1885	·1109
15	·0717	·0966	·1128	·0718
20	·0531	·0704	·0805	·0531
25	·0422	·0554	·0626	·0421
30	·0351	·0456	·0512	·0349
35	·0300	·0389	·0433	·0298

5. ACKNOWLEDGMENT

The author is grateful to Dr. Jiří Anděl for his guidance of my diploma work which was a basis of the present paper.

References

[1] *Hájek, J.* and *Anděl, J.*: Stationary processes (Lecture Notes). SPN, Praha, 1969.
[2] *Hurt, J.*: Tests for the goodness of fit in stationary autoregressive sequences. (Unpublished thesis.)
[3] *Ibragimov, I. A.* and *Linnik, Yu. V.*: Independent and stationary sequences of random variables (Russian), Nauka, Moskva, 1965.
[4] *Kendall, M. G.*: Tables of autoregressive series. Biometrika 36 (1949), p. 267.
[5] *Loève, M.*: Probability theory (3d ed.). D. Van Nostrand, New York, 1963.
[6] *Plackett, R. L.*: A reduction formula for normal multivariate integrals. Biometrika 41 (1954), p. 351.
[7] *Rozanov, Yu. V.*: Stationary random processes (Russian). GIFML, Moskva, 1963.
[8] *Wold, H.*: Random normal deviates. Tracts for Computers, No. XXV, Cambridge, 1948.

Souhrn

O JEDNODUCHÉM ODHADU KORELAČNÍ FUNKCE STACIONÁRNÍCH NÁHODNÝCH POSLOUPNOSTÍ

JAN HURT

Nechť $\{X_t\}_{t\in T}$ je stacionární gaussovský diskrétní proces, kde T je množina celých čísel. Předpokládejme $ET_t = 0$, $t \in T$, a označme $Z_t = \operatorname{sign} X_t$, $T_{tj} = Z_t Z_{t+j}$ pro j přirozené. Ukazuje se, že $ET_{tj} = 2 \arcsin \varrho_j/\pi$, kde ϱ_j je korelační koeficient X_t a X_{t+j}, takže veličiny T_{tj} mohou být užity k odhadu korelační funkce původního procesu $\{X_t\}_{t\in T}$. Dále je odvozen vzorec pro výpočet cov (T_{0j}, T_{kj}). Asymptotické vlastnosti průměru $\bar{T}_j = \sum_{t=1}^{N-j} T_{tj}/(N-j)$ jsou studovány za předpokladu, že spektrální hustota procesu $\{X_t\}_{t\in T}$ je nenulová a má ohraničenou druhou derivaci. Speciálně odvozené výsledky platí pro stacionární gaussovské autoregresní posloupnosti, což je prakticky nejdůležitější případ. Je dokázáno, že $\bar{T}_j$ je asymptoticky normální, a že posloupnost $\{T_{tj}\}_{t\in T}$ splňuje zákon velkých čísel. Na závěr jsou uvedeny některé numerické příklady a Monte-Carlo studie navrhované statistiky.

Author's address: RNDr. *Jan Hurt*, Matematicko-fyzikální fakulta University Karlovy, Sokolovská 83, 186 00 Praha 8.