

AUTOKORELOGRAM ZNAKOWY

Niniejsza praca powstała w Dziale statystycznej kontroli jakości Instytutu Matematycznego PAN jako zespołowa. Koncepcję autokorelogramu znakowego (§§ 1, 2) podał w roku 1955 J. Oderfeld, który ponadto zbadał pewien fragment jego teorii (§ 6). Całość teorii (§§ 3, 4, 5, 7) opracował w latach 1955 i 1956 P. Mikulski. Część eksperymentalną (§§ 8, 9, 10) wykonał W. Rudzki⁽¹⁾.

§ 1. W wielu badaniach eksperymentalnych występują tzw. *serie czasowe*, to znaczy ciągi liczb wyrażające zależność pewnej wielkości od czasu. Poprzestaniemy na podaniu paru przykładów takich wielkości: ciśnienie barometryczne, zawartość CO₂ w spalinach z kotła, cena pszenicy, stan wody w rzece, frakcja braków w produkcji papierosów itd. Na ogół czas bywa wyrażany w zwykłych jednostkach: minutach, godzinach, dniach, jednakże nieraz dogodnie jest przyjąć jednostkę umowną; na przykład serią czasową nazwiemy również ciąg liczb wyrażających zużycie opony po przejechaniu tysiąca, dwóch tysięcy itd. kilometrów.

Ważną cechą serii czasowej jest wzajemna zależność jej wyrazów. Badanie jej pozwala często wykryć pewną tendencję ogólną, na przykład przybór wody w rzece, pogarszanie się jakości produkcji itd.

Powszechnie stosowaną miarą tej współzależności jest tzw. *współczynnik $r(k)$ autokorelacji* zdefiniowany w następujący sposób:

Niechaj x_1, x_2, \dots, x_n będzie ciągiem liczb rzeczywistych. Z definicji

$$(1) \quad r(k) = \frac{\frac{1}{n-k} \sum_{j=1}^{n-k} x_j x_{j+k} - \frac{1}{(n-k)^2} \left(\sum_{j=1}^{n-k} x_j \right) \left(\sum_{j=1}^{n-k} x_{j+k} \right)}{\left[\left(\frac{1}{n-k} \sum_{j=1}^{n-k} x_j^2 - \frac{1}{(n-k)^2} \left(\sum_{j=1}^{n-k} x_j \right)^2 \right) \left[\frac{1}{n-k} \sum_{j=1}^{n-k} x_{j+k}^2 - \frac{1}{(n-k)^2} \left(\sum_{j=1}^{n-k} x_{j+k} \right)^2 \right] \right]^{1/2}}$$

⁽¹⁾ Po zgłoszeniu pracy do druku zapoznaliśmy się z artykułem [2], w którym jest rozwinięta myśl podobna. Z. A. Łomnicki i S. K. Zaremba tworzą zastępczy proces zero-jedynkowy i szacują jego współczynniki autokorelacji klasycznej, co prowadzi dokładnie do naszego wzoru (2). Na podstawie pracy [2] można natychmiast otrzymać wyniki, które podajemy w § 4 i § 7 naszego artykułu.

Obliczanie $r(k)$ nie nastręcza trudności koncepcyjnych, jednakże w długich seriach występują trudności rachunkowe, które nawet zmusiły do stosowania specjalnych maszyn, autokorelatorów, przeważnie elektrycznych.

W niniejszej pracy definiujemy prostszy współczynnik autokorelacji, mianowicie współczynnik korelacji znakowej, badamy jego niektóre własności i podajemy przykłady zastosowania.

§ 2. Niechaj znowu będzie ciąg liczb rzeczywistych x_1, x_2, \dots, x_n będących realizacjami ciągu zmiennych losowych $\{X_i\}$. Oznaczmy przez $E(X_i)$ ich wartości oczekiwane oraz przez y_i różnicę $x_i - E(X_i)$. Tworzymy ciąg znaków $\text{sign } y_1, \text{sign } y_2, \dots, \text{sign } y_n$, a następnie ciąg $\text{sign}(y_1 y_{1+k}), \text{sign}(y_2 y_{2+k}), \dots, \text{sign}(y_{n-k} y_n)$. Niechaj w tym ciągu będzie z_k znaków „plus”. Wyrażenie

$$(2) \quad \varrho(k) = 2z_k / (n - k) - 1$$

nazywamy *współczynnikiem* autokorelacji znakowej serii skończonej rzędu k . Ponieważ $0 \leq z_k \leq n - k$, więc $-1 \leq \varrho(k) < 1$.

W dalszych rozdziałach pokażemy, że zachodzi duża analogia między własnościami $\varrho(k)$ a $r(k)$, a przy pewnych założeniach nawet zupełna zgodność. Natomiast obliczanie $\varrho(k)$ jest dużo prostsze niż obliczanie $r(k)$ i nadaje się do zautomatyzowania najprostszymi środkami, o których będzie mowa później (§ 8).

§ 3. Jednym z podstawowych zagadnień analizy serii czasowych jest weryfikacja hipotez dotyczących typu serii. Zwykle przyjmuje się cztery podstawowe typy:

1. Serie czysto przypadkowe.
2. Serie średniej ruchomej.
3. Serie harmoniczne.
4. Serie autoregresyjne.

U podstaw tej umownej klasyfikacji leży twierdzenie Wolda o dekompozycji procesów (zob. [3]). Klasyfikacja ta jest dość powszechnie przyjęta, nie będziemy się więc nad nią dłużej zatrzymywali. Zwrócimy tylko uwagę, że wymienione klasy nie są rozłączne i że np. klasa 4 obejmuje jako szczególne przypadki klasy 1 i 2. W dalszym ciągu zajmiemy się porównaniem oczekiwanego przebiegu funkcji autokorelacji znakowej z klasyczną funkcją autokorelacyjną.

§ 4. Rozpatrzmy przede wszystkim serię typu 2.

Niechaj rozpatrywana seria czasowa będzie utworzona przez zastosowanie metody średniej ruchomej o wagach a_i ($i = 1, 2, \dots, m$) do ciągu

niezależnych zmiennych losowych ε_j o jednakowych rozkładach. Znaczący to, że wyrazy serii są realizacjami ciągu zmiennych losowych

$$(3) \quad Y_j = \sum_{i=1}^m a_i \varepsilon_{j+i-1}.$$

Jak wiadomo [1], funkcja autokorelacyjna jest w tym przypadku następująca:

$$(4) \quad \bar{r}(k) = \begin{cases} \left(\sum_{i=1}^{m-k} a_i a_{i+k} \right) / \sum_{i=1}^m a_i^2 & \text{dla } k < m, \\ 0 & \text{dla } k \geq m. \end{cases}$$

Jeżeli przyjmiemy wagi jednakowe

$$(5) \quad a_i = 1/m,$$

to

$$(6) \quad \bar{r}(k) = \begin{cases} 1 - k/m & \text{dla } k < m, \\ 0 & \text{dla } k \geq m. \end{cases}$$

Wyznamy dla tego samego przypadku wartość oczekiwaną $\bar{q}(k)$ współczynnika autokorelacji znakowej.

Najpierw określimy zmienne losowe $Y_j^{(k)}$ i Z_k :

$$(7) \quad Y_j^{(k)} = \begin{cases} 0, & \text{gdy } \text{sign}(y_j y_{j+k}) = -, \\ 1, & \text{gdy } \text{sign}(y_j y_{j+k}) = +. \end{cases}$$

$$(8) \quad Z_k = \sum_{j=1}^{n-k} Y_j^{(k)}.$$

Teraz założymy, że zmienne losowe ε_j mają łączny rozkład normalny. Jest to ograniczenie ogólności, ale niezbyt dotkliwie. Drugie założenie, że $E(\varepsilon_j) = 0$, a więc również $E(Y_j) = 0$, nie zmniejsza ogólności dalszych rozważań i jest wprowadzone tylko dla uproszczenia.

Przy tych założeniach łączny, n -wymiarowy rozkład zmiennych losowych ε_j ma środek symetrii w początku układu współrzędnych.

Oznaczmy łączną n -wymiarową dystrybuantę zmiennych losowych ε_j przez $F(\varepsilon)$. Wobec tego

$$(9) \quad P(Y_j^{(k)} = 1) = 1 - P(Y_j^{(k)} = 0) = 1 - P(y_j y_{j+k} < 0) = \\ = 1 - 2 \int \dots \int_{(y_j < 0, y_{j+k} > 0)} dF(\varepsilon).$$

Ze względu na założoną normalność rozkładu, prawdopodobieństwo $P(y_j y_{j+k} < 0)$ jest proporcjonalne do kąta między hiperpłaszczyznami



$y_j = 0$ i $y_{j+k} = 0$. Oznaczmy ten kąt literą φ . Jeśli teraz oznaczmy

$$(10) \quad A = \left(\sum_{i=1}^{m-k} a_i a_{i+k} \right) / \sum_{i=1}^m a_i^2,$$

to będzie, jak wiadomo,

$$(11) \quad \varphi = \arccos A$$

przy dodatkowym warunku, że $0 \leq \varphi \leq \pi$. Stąd

$$(12) \quad P(Y_j^{(k)} = 1) = 1 - \frac{2\varphi}{2\pi} = 1 - \frac{1}{\pi} \arccos A.$$

Z definicji zmiennych $Y_j^{(k)}$ wynika, że $E(Y_j^{(k)}) = P(Y_j^{(k)} = 1)$. Ponieważ zaś z założenia $E(Y_j^{(k)})$ ma tę samą wartość dla wszelkich j , więc wobec (8)

$$E(Z_k) = (n-k)E(Y_j^{(k)}) = (n-k)P(Y_j^{(k)} = 1).$$

Uwzględniając jeszcze wzór (2) znajdujemy, że

$$(13) \quad \bar{\varrho}(k) = E(\varrho(k)) = E(2Z_k/(n-k)-1) = 2P(Y_j^{(k)} = 1) - 1.$$

Z (13) i (12) otrzymujemy

$$\bar{\varrho}(k) = \begin{cases} 1 - \frac{2}{\pi} \arccos A & \text{dla } k < m, \\ 0 & \text{dla } k \geq m. \end{cases}$$

Można to napisać również w postaci

$$(14) \quad \bar{\varrho}(k) = \begin{cases} \frac{2}{\pi} \arcsin A & \text{dla } k < m, \\ 0 & \text{dla } k \geq m, \end{cases}$$

przy dodatkowym warunku, że $-1 \leq \bar{\varrho}(k) \leq 1$. Warunek ten jest potrzebny, bo wyrażenie $\arcsin a$ nie jest jednoznaczna funkcją a . Równość $\bar{\varrho}(k) = 0$ dla $k \geq m$ jest oczywista, ponieważ w tym przypadku zmienne losowe Y_j oraz Y_{j+k} są niezależne, co wobec symetrii daje $P(Y_j^{(k)} = 1) = \frac{1}{2}$. W szczególnym przypadku, gdy $a_i = 1/m$,

$$(15) \quad \bar{\varrho}(k) = \frac{2 \arcsin(1-k/m)}{\pi} \quad \text{dla } k < m.$$

Zauważmy teraz, że seria czysto przypadkowa (typ 1) jest szczególnym przypadkiem rozważanej serii, mianowicie takim, że $m = 1$ i $a_i = 1$. Wówczas

$$(16) \quad \bar{\varrho}(k) = \begin{cases} 1 & \text{dla } k < 1, \\ 0 & \text{dla } k \geq 1. \end{cases}$$

Zbierając dotychczasowe wyniki spostrzegamy, że w seriach typu 2 (a również typu 1) funkcje $\bar{r}(k)$ i $\bar{\varrho}(k)$ mają podobne własności:

a. Długość odcinka osi k , dla którego zarówno $\bar{r}(k)$, jak $\bar{\varrho}(k)$ są różne od zera, jest ta sama. Znając doświadczalny przebieg $r(k)$ lub $\varrho(k)$ można na tej podstawie wyciągnąć wniosek co do liczby m elementów średniej ruchomej.

b. Jeżeli seria jest dostatecznie długa i $\varrho(k)$ można traktować jako dobre przybliżenie $\bar{\varrho}(k)$, to stosując transformację

$$\bar{r}(k) = \sin[\bar{\varrho}(k)\frac{1}{2}\pi]$$

można uzyskać dobre oszacowanie $\bar{r}(k)$.

c. W seriach średnich ruchomych różnice liczbowe między wartościami wynikającymi z wzorów (4) i (14) nie przekraczają 0,21. Wynika to z własności funkcji $d = (z/\pi)\arcsin A - A$.

§ 5. Rozpatrzmy teraz przypadek tak zwanej *serii harmonicznej* (typ 3 według naszej klasyfikacji).

Niechaj wyrazy serii będą realizacjami ciągu zmiennych losowych $\{Y_j\}$ takiego, że

$$(17) \quad Y_j = D \sin \theta j + \varepsilon_j,$$

przy czym $\theta = 2\pi a/b$, gdzie a i b są liczbami naturalnymi, względnie pierwszymi, a ε_j jest ciągiem zmiennych losowych niezależnych o jednakowych rozkładach.

Dla porządku notujemy, że jest to przypadek szczególny stacjonarnego procesu stochastycznego harmonicznego, w którym składnik okresowy jest skończoną sumą wyrazów harmonicznych (zob. np. [4]). Przyjęliśmy tylko jeden wyraz, analogicznie jak w pracy [1] (str. 406), gdyż uogólnienie nie wprowadziłoby — jak sądzimy — istotnych różnic w metodzie badania i w wynikach.

Klasyczna funkcja autokorelacyjna dla ustalonego k jest w tym przypadku (zob. [1]) następująca:

$$(18) \quad \bar{r}(k) = C \cos \theta k,$$

gdzie C oznacza stałą. Jest to więc funkcja okresowa o okresie b . Wykażemy dalej, że oczekiwany przebieg funkcji autokorelacji znakowej ma podobną właściwość, to znaczy jest to w przybliżeniu funkcja okresowa o okresie b .

Niech $\{\varepsilon_j\}$ będzie ciągiem, tak jak w przypadku poprzednim, zmiennych losowych niezależnych o jednakowych rozkładach i wartości oczekiwanej równej zeru. Przyjmijmy dla uproszczenia $D = 1$. Przy tych zało-

zeniach i przy określeniu (7) zachodzi zależność

$$\begin{aligned} P(Y_j^{(k)} = 1) &= 1 - P(Y_j < 0; Y_{j+k} > 0) - P(Y_j > 0; Y_{j+k} < 0) = \\ &= 1 - P(\sin\theta j + \varepsilon_j < 0; \sin\theta(j+k) + \varepsilon_{j+k} > 0) - \\ &\quad - P(\sin\theta j + \varepsilon_j > 0; \sin\theta(j+k) + \varepsilon_{j+k} < 0) = \\ &= 1 - P(\varepsilon_j < -\sin\theta j; \varepsilon_{j+k} > -\sin\theta(j+k)) - \\ &\quad - P(\varepsilon_j > -\sin\theta j; \varepsilon_{j+k} < -\sin\theta(j+k)). \end{aligned}$$

Wobec założonej niezależności zmiennych losowych ε_j znajdujemy dalej, oznaczając przez $F(\dots)$ dystrybuantę zmiennej losowej ε_j ,

$$\begin{aligned} P(Y_j^{(k)} = 1) &= 1 - F(-\sin\theta j) \{1 - F(-\sin\theta(j+k))\} - \\ &\quad - (1 - F(-\sin\theta j)) F(-\sin\theta(j+k)), \end{aligned}$$

a po uporządkowaniu

$$(19) \quad P(Y_j^{(k)} = 1) = 1 - F(-\sin\theta j) - F[-\sin\theta(j+k)][1 - 2F(-\sin\theta j)].$$

Zatem zgodnie z określeniami (8) i (2)

$$\bar{\varrho}(k) = E\left(2 \frac{z_k}{n-k} - 1\right) = 2E\left(\frac{z_k}{n-k}\right) - 1 = 2E\left(\frac{1}{n-k} \sum_{j=1}^{n-k} Y_j^{(k)}\right) - 1,$$

czyli

$$(20) \quad \begin{aligned} \bar{\varrho}(k) &= 1 - \frac{2}{n-k} \sum_{j=1}^{n-k} F(-\sin\theta j) - \\ &\quad - \frac{2}{n-k} \sum_{j=1}^{n-k} F[-\sin\theta(j+k)][1 - 2F(-\sin\theta j)]. \end{aligned}$$

Funkcja (20) składa się z trzech członów: pierwszego stałego, drugiego równego podwójnej średniej arytmetycznej funkcji $F(-\sin\theta j)$ na odcinku $(1, n-k)$ i wreszcie trzeciego, równego podwójnej średniej arytmetycznej funkcji $F[-\sin\theta(j+k)][1 - 2F(-\sin\theta j)]$ na odcinku $(1, n-k)$. Człon drugi jest funkcją okresową argumentu k o okresie b , gdyż średnia arytmetyczna funkcji okresowych na ustalonym odcinku jest funkcją okresową długości tego odcinka.

Wreszcie można wykazać (zrobimy to w § 6, aby nie przerywać ciągłości), że trzeci człon jest w przybliżeniu okresowy z okresem b . Przybliżenie to jest dobre, gdy n jest duże w porównaniu z k , co zawsze zachodzi w praktyce.

Z tego wynika, że $\bar{\varrho}(k)$ jest z dobrym przybliżeniem funkcją okresową o okresie b , czyli że ma tę samą własność, którą — w tym przypadku ściśle — posiada $\bar{r}(k)$.

Wnioski te są ważne nie tylko dla serii określonej przez (17), ale dla wszelkiej serii typu $Y_j = f(j) + \varepsilon_j$, gdzie $f(j)$ jest dowolną funkcją okresową.

§ 6. Zgodnie z zapowiedzią przy końcu poprzedniego paragrafu, zbadamy własności trzeciego członu wyrażenia (20). Oznaczmy

$$(21) \quad \begin{aligned} S(k) &= \sum_{j=1}^{n-k} F[-\sin\theta(j+k)][1-2F(-\sin\theta_j)], \\ s(k) &= \frac{2}{n-k} S(k), \end{aligned}$$

gdzie $k \leq n$, $F(\dots)$ jest dowolną dystrybuantą, $\theta = 2\pi a/b$, przy czym a i b są naturalne i względnie pierwsze.

TWIERDZENIE 1. *Funkcja $S(k)$ określona przez (21) jest sumą funkcji okresowej gasnącej skokowo^(*) i funkcji okresowej.*

Dla dowodu oznaczmy

$$(22) \quad \left\{ \begin{aligned} F_j &= F(-\sin\theta_j) = F\left(-\sin 2\pi \frac{a}{b} j\right), \\ A_{jk} &= F_{j+k}(1-2F_j), \\ B_k &= \sum_{j=1}^b A_{jk}. \end{aligned} \right.$$

Łatwo sprawdzamy, że funkcje F_j , A_{jk} , B_k są okresowe o okresie b . Dokładniej mówiąc, zachodzą relacje

$$F_{j+b} = F_j, \quad A_{j+b,k} = A_{jk}, \quad B_{k+b} = B_k.$$

Napiszmy $n-k$ w postaci

$$(23) \quad n-k = N_k b + r, \quad ,$$

gdzie $N_k = [(n-k)/b]$. Reszta r może więc przybierać wartości $0, 1, 2, \dots, b-1$ i jest funkcją okresową zmiennej k o okresie b .

Wobec (21), (22) i (23) można napisać

$$S_k = \sum_{j=1}^{N_k b} A_{jk} + \sum_{j=N_k b+1}^{n-k} A_{jk},$$

co z uwagi na okresowość A_{jk} daje

$$(24) \quad S_k = N_k B_k + \sum_{j=1}^r A_{jk}.$$

(*) Przez to pojęcie rozumiemy iloczyn funkcji okresowej o okresie T i nierosnącej funkcji o przedziałach stałości o długości T .

Wyrażenie $\sum_{j=1}^r A_{jk}$ jest, podobnie jak B_k , funkcją okresową zmiennej k o okresie b . Jeżeli więc k przebiega wartości $0, 1, 2, \dots, n$, to N_k maleje skokami od $[n/b]$ do zera. A więc $S(k)$ jest sumą funkcji okresowej gasnącej skokowo i funkcji okresowej, co stanowi tezę twierdzenia 1.

TWIERDZENIE 2. *Jeśli dla ustalonego k , $n \rightarrow \infty$, to funkcja $s(k)$ określona przez (21) dąży do funkcji okresowej o okresie b .*

Dla dowodu zwrócimy uwagę, że z określenia A_{jk} za pomocą wzoru (22) wynika, że dla każdej dystrybuanty $F(\dots)$ oraz każdego b można znaleźć takie dwie liczby d i g niezależne od n i od k , że zachodzi relacja

$$(25) \quad d < \sum_{j=1}^r A_{jk} < g.$$

Z (24) i (25) wynika, że

$$N_k B_k + d < S(k) < N_k B_k + g,$$

a wobec (21)

$$(26) \quad \frac{2N_k}{n-k} B_k + \frac{2d}{n-k} < s(k) < \frac{2N_k}{n-k} B_k + \frac{2g}{n-k}.$$

Jeżeli $n \rightarrow \infty$, to $n-k \rightarrow bN_k$, $2d/(n-k) \rightarrow 0$ oraz $2g/(n-k) \rightarrow 0$, wobec czego

$$(27) \quad s(k) \rightarrow \frac{2}{b} B_k,$$

co kończy dowód.

Zwróćmy wreszcie uwagę, że do dowodów twierdzeń 1 i 2 nie korzystaliśmy z żadnych innych własności funkcji sinus w określeniu (17) poza jej okresowością. A więc twierdzenia 1 i 2 pozostają prawdziwe, gdy w określeniu (21) zastąpimy funkcję sinus przez jakąkolwiek funkcję okresową.

§ 7. Omówmy teraz typ 4 serii czasowych, czyli tak zwane *serie autoregresyjne*. W ogólnym przypadku taka seria jest zdefiniowana przez równanie

$$(28) \quad Y_j = f(Y_{j-1}, Y_{j-2}, \dots, Y_{j-s}) + \varepsilon_j,$$

gdzie $\{\varepsilon_j\}$ jest ciągiem zmiennych losowych niezależnych o jednakowych rozkładach. Zwykle rozpatruje się prosty przypadek równania liniowego

$$(29) \quad Y_j = -a_1 Y_{j-1} - a_2 Y_{j-2} - \dots - a_s Y_{j-s} + \varepsilon_j.$$

Takie równanie nosi nazwę *równania autoregresji* rzędu s . Ograniczymy się przykładowo do rozważenia równania autoregresji drugiego rzędu. Jak wiadomo [1], w tym przypadku rozwiązanie równania różnicowego (29) sprowadza się do postaci

$$(30) \quad Y_j = p^j (A \cos \theta j + B \sin \theta j) + \sum_{i=0}^{\infty} \xi_i \varepsilon_{j-i+1},$$

gdzie

$$\cos \theta = -\frac{a_1}{2\sqrt{a_2}}, \quad p = \sqrt{a_2}, \quad \xi_i = \frac{2}{\sqrt{4a_2 - a_1^2}} p^i \sin \theta i,$$

przy dodatkowych warunkach $0 < a_2 < 1$ oraz $4a_2 - a_1^2 > 0$. Te warunki mają znaczenie praktyczne, w szczególności warunek $a_2 < 1$, a więc $p < 1$, w przeciwnym bowiem przypadku ze wzrostem j wahania Y_j rosłyby nieograniczenie. Jak widać, Y_j składa się z dwóch członów: pierwszego harmonicznego z tłumionymi wagami oraz drugiego — średniej ruchomej z nieskończoną liczbą składników o wagach ξ_i . Przesuwając początek układu możemy dowolnie zmniejszyć człon pierwszy i rozpatrywać jedynie serię

$$(31) \quad Y_j = \sum_{i=0}^{\infty} \xi_i \varepsilon_{j-i+1}.$$

Dla serii średniej ruchomej o skończonej liczbie wyrazów zachodzi (por. § 4) następująca zależność:

$$(32) \quad \bar{\varrho}(k) = \frac{2}{\pi} \arcsin \bar{r}(k).$$

Dla rozpatrywanej serii klasyczna funkcja autokorelacyjna jest (zob. [1]) postaci

$$(33) \quad \bar{r}(k) = p^k \frac{\sin(k\theta + \Psi)}{\sin \Psi}, \quad \text{gdzie} \quad \operatorname{tg} \Psi = \frac{1+p^2}{1-p^2} \operatorname{tg} \theta.$$

Przyjmując prawdziwość relacji (32) dla średniej z nieskończonej liczby elementów, znajdujemy z (22) i (23)

$$(34) \quad \bar{\varrho}(k) = \frac{2}{\pi} \arcsin \left[p^k \frac{\sin k\theta + \Psi}{\sin \Psi} \right].$$

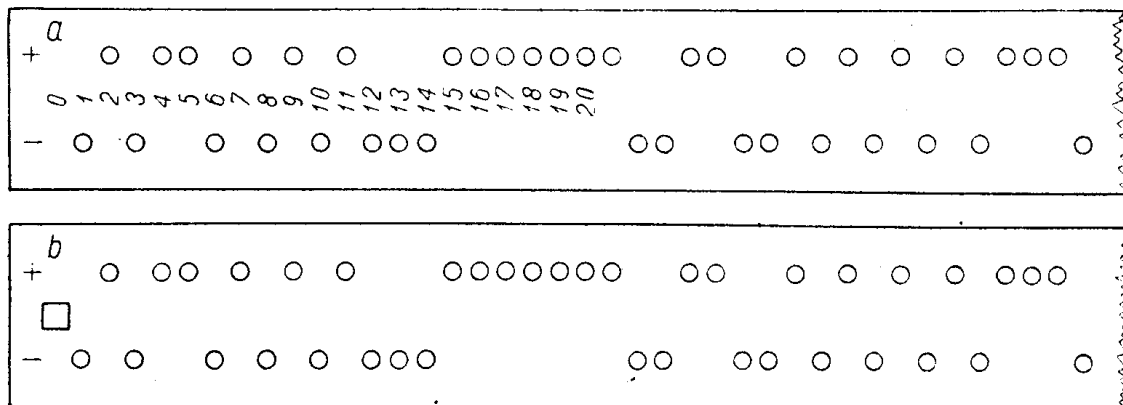
Porównanie (33) i (34) wykazuje, że $\bar{\varrho}(k)$ i $\bar{r}(k)$ mają podobne własności: są okresowymi funkcjami gasnącymi o tym samym okresie i o podobnym sposobie tłumienia amplitudy (czynnik p^k). Podobnie jak

to było w przypadku serii średnich ruchomych (por. § 4) różnica między $\bar{r}(k)$ a $\bar{\rho}(k)$ nie przekracza 0,21.

§ 8. Technika obliczania współczynnika $\rho(k)$ korelacji znakowej jest bardzo prosta. Po scentralizowaniu serii otrzymanej z doświadczenia przez odjęcie od wyrazów serii oryginalnej ich średniej, tworzy się odpowiadającą serię znaków plus i minus, przy czym znak plus (minus) stawia się, jeśli wyraz serii scentralizowanej był większy (mniejszy) od średniej.

Dogodnie jest tę serię znaków napisać w równych odstępach przez kalkę na dwóch paskach papieru. Teraz ustawiamy paski koło siebie z przesunięciem o k znaków i liczymy, w ilu miejscach znaki na paskach są zgodne. W ten sposób otrzymujemy liczbę z_k i stosujemy wzór (2).

Można również zarejestrować serię znaków w postaci odpowiednich otworków jednakowo rozmieszczonych na dwóch paskach (rys. 1). Znak plus zaznaczamy otworkiem nad osią, znak minus pod osią. Pasek a ma dodatkową skalę numerów k , a pasek b ma okienko. Aby znaleźć z_k , wystarczy położyć pasek b na pasku a tak, żeby w okienku był widoczny numer k i policzyć otworki przeświecające przez oba paski.



ZM-223

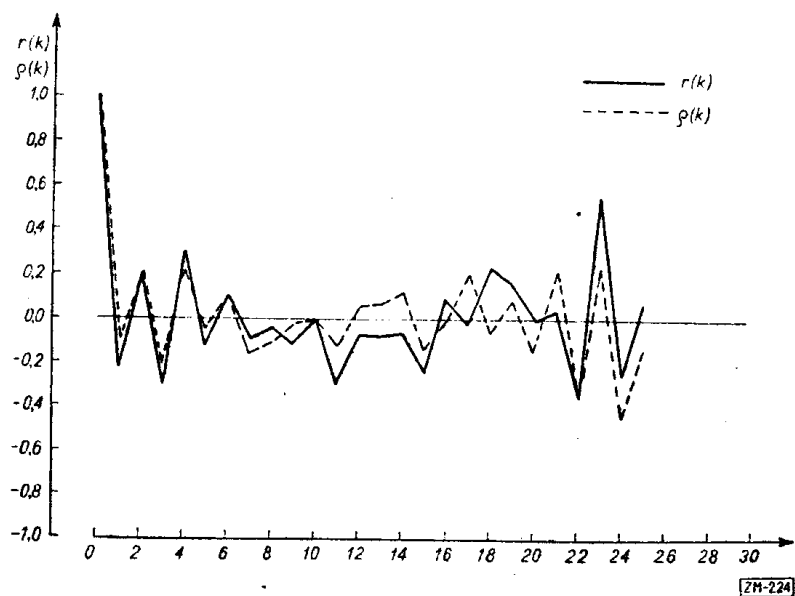
Rys. 1

Obliczanie $\rho(k)$ jest bardzo szybkie i łatwe oraz daje mało okazji do pomyłek; można je powierzyć personelowi, który nie zdołałby obliczyć $r(k)$. Oczywiście interpretacja autokorelogramów jest w obu przypadkach niełatwa i odpowiedzialna.

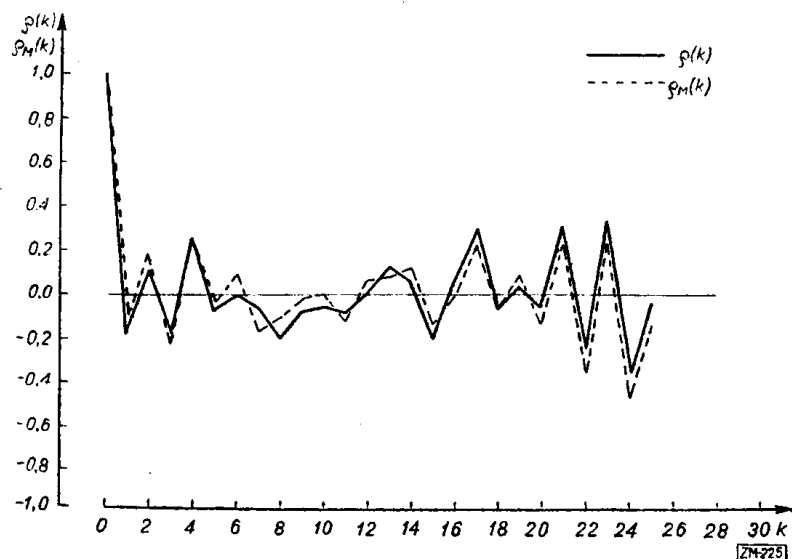
§ 9. Aby zilustrować podobieństwo korelogramów $\rho(k)$ i $r(k)$ wyjaśnione w poprzednich paragrafach, zrobiliśmy kilka doświadczeń.

Najpierw zbudowano serię liczb przypadkowych, składającą się z 50 elementów pobranych z tablicy liczb przypadkowych. Autokorelogram tej serii wyznaczony za pomocą obu współczynników jest na

rysunku 2. Są tu pewne różnice w poszczególnych punktach wykresu, jednak charakter przebiegu obu autokorelogramów jest zgodny i w obu przypadkach wskazuje przypadkowość badanej serii. O zgodności świadczy również współczynnik korelacji klasycznej między obu autokorelogramami wynoszący 0,723.



Rys. 2

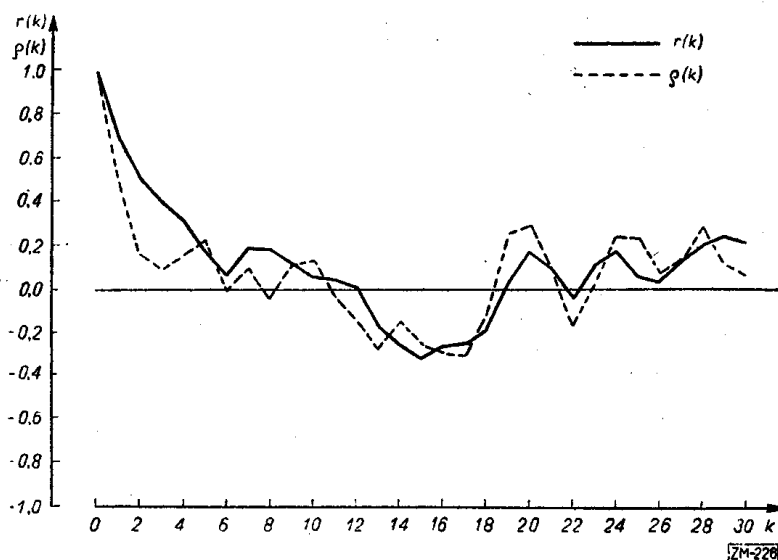


Rys. 3

Na rysunku 3 powtórzono autokorelogram znakowy $\rho(k)$ z rysunku 2 i dodano autokorelogram znakowy $\rho_M(k)$ tej samej serii wyznaczo-

ny jak poprzednio, z tą jednak różnicą, że centralizowano względem mediany. Jak widzimy, użycie mediany nie zmieniło charakteru przebiegu autokorelogramu. Jest to informacja o tyle cenna, że w pewnych przypadkach łatwiej jest wyznaczyć medianę szeregu niż średnią, pozwala to więc na dalsze uproszczenie pracy obliczeniowej.

Na rysunku 4 podano autokorelogram klasyczny i znakowy serii czasowej złożonej z 54 wyrazów i przedstawiającej wyniki prób wytrzymałości cementu wyprodukowanego w pewnym okresie czasu. I w tym przypadku, jak wynika z wykresu, charakter obu autokorelogramów jest zgodny. Wprawdzie różnice w poszczególnych punktach są tu większe niż w poprzednim przykładzie, jednak charakter przebiegu jest bardziej zgodny, o czym świadczy współczynnik korelacji wynoszący 0,801.



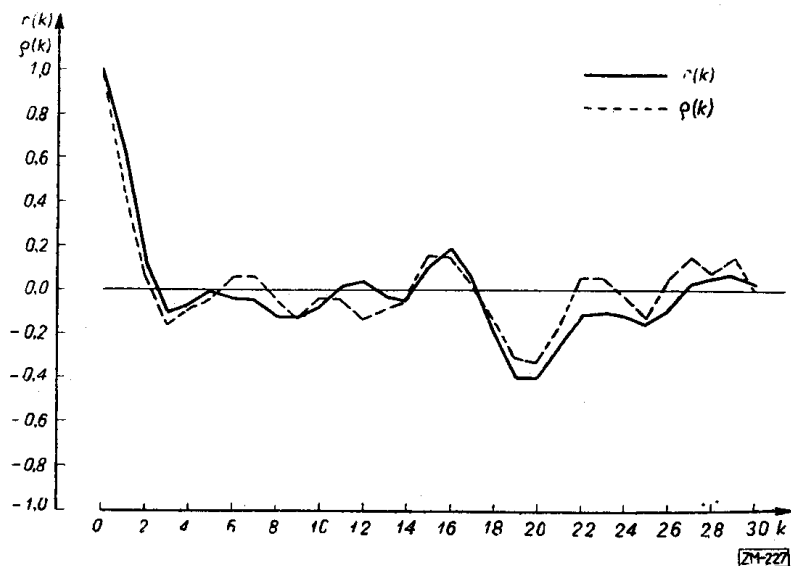
Rys. 4

W trzecim przypadku porównanie przeprowadzono na serii czasowej, przedstawiającej średnie roczne ceny pszenicy w okresie stu kolejnych lat (rys. 5). Dane liczbowe przyjęto według pracy [3]. Zgodność autokorelogramów jest tu jeszcze lepsza niż w obu poprzednich przypadkach; współczynnik korelacji wynosi w tym przypadku 0,877.

Porównanie współczynników autokorelacji przeprowadzone na powyższych przykładach wskazuje, że zastąpienie metody klasycznej metodą znakową, znacznie prostszą i łatwiejszą w obliczeniach, jest możliwe i nie wpływa poważnie na ostateczne wnioski. Należy tu zwrócić uwagę, że wnioskowanie z autokorelogramów, zarówno klasycznych jak znakowych, jest ryzykowne, gdy seria jest krótka (n małe).

Podczas prac w Łódzkiej Wytwórni Papierosów (w roku 1956) nad polepszeniem jakości papierosów metodami statystycznymi zajęto się

między innymi kontrolą średnicy papierosa. W czasie badań zauważono, że mimo dokładnego uregulowania maszyny papierosowej wyprodukowane papierosy przy zachowaniu innych własności na odpowiednim poziomie mają różne średnice, co odbija się niekorzystnie na tzw. nabiciu papierosa. Dalsza regulacja maszyny nie usunęła tych różnic.



Rys. 5

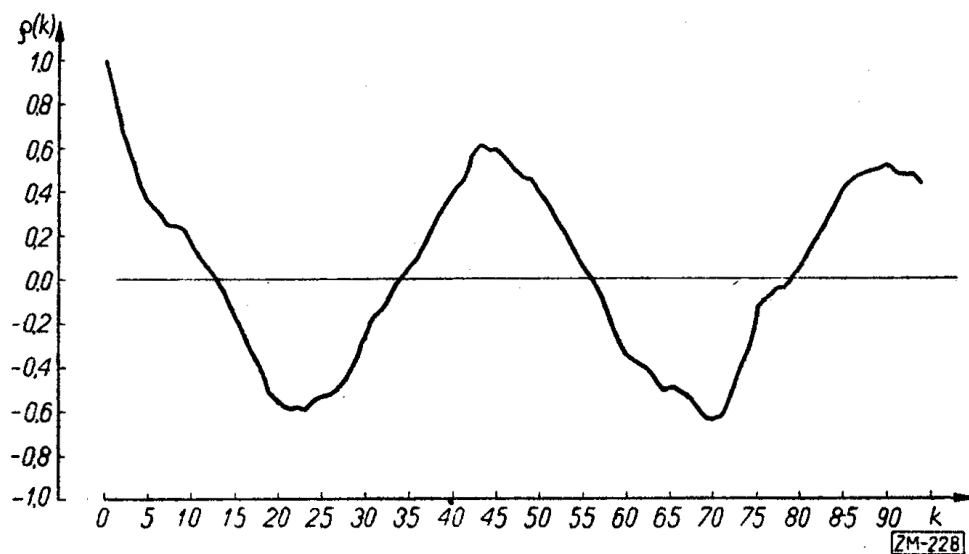
Nasunęło się wówczas przypuszczenie, że zmiany w średnicy mogą powstawać wskutek wad taśmy formującej wałek papierosowy, który następnie tnie się na papierosy. Jeden obieg taśmy w koło trwa około 2 sekund i daje 44 papierosy. Gdyby przypuszczenie było słuszne, wówczas zmiany średnicy musiałyby występować cyklicznie. Aby to wyjaśnić, postanowiono zastosować metodę autokorelacji znakowej.

Ponieważ ze względów technicznych niemożliwe jest pobieranie z maszyny papierosów w kolejności ich wyprodukowania, pobrano 93 papierosy, biorąc co drugi kolejny papieros. Zmierzono średnicę każdego papierosa w obu jego końcach. W ten sposób otrzymano 186 pomiarów średnicy w kolejnych punktach przecięcia wałka.

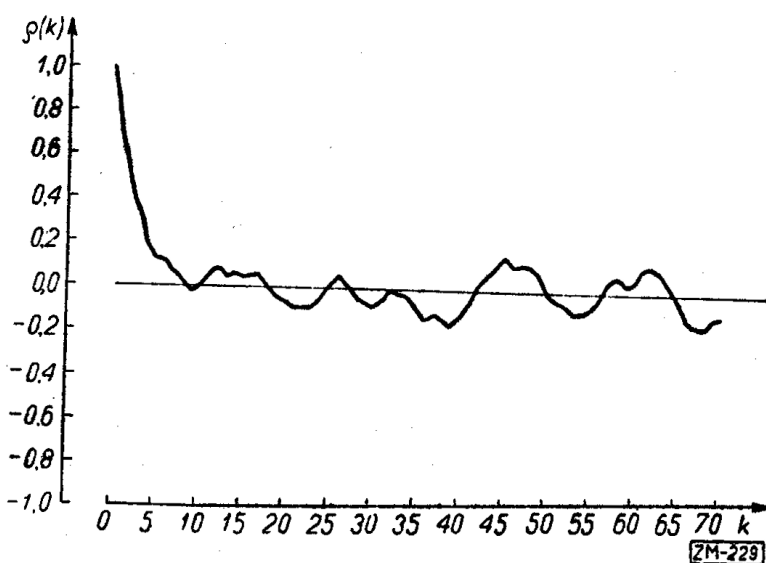
Aby zmniejszyć wpływ przyczyn ubocznych, wyrównano otrzymany szereg czasowy, obliczając średnią ruchomą z 5 elementów o jednakowych wagach. Następnie obliczono wartości kolejnych współczynników autokorelacji od 1 do 94 i sporządzono autokorelogram (rys. 6).

Uzyskany wynik wskazuje wyraźnie istnienie cyklicznych zmian średnicy. Długość cyklu wynosi 43-45 papierosów. Ponieważ długość ta pokrywa się z ilością papierosów mieszczących się na obwodzie taśmy papierosowej (około 44 papierosów), więc hipoteza, że powodem była wadliwa taśma, uzyskała mocne potwierdzenie.

W następnym doświadczeniu z inną taśmą uzyskano nowy ciąg 188 pomiarów. Odpowiedni autokorelogram pokazano na rysunku 7. Nie widać tu wyraźnej cykliczności, którą można by przypisać wadom taśmy.



Rys. 6



Rys. 7

Dalsze doświadczenia dawały korelogramy podobne albo do rysunku 6, albo do rysunku 7.

Ostateczny wniosek był taki, że zdarzają się taśmy z pewnym typem wad wpływających w sposób znany i niekorzystny na własności użytkowe papierosów. Pokazało się następnie, że wady te są ukryte, czyli że nie można środkami badania dostępnymi obecnie odróżnić taśmy

добрей od niedobrej przed jej założeniem do maszyny papierosowej. Przyпускаjemy, że taśmy pochodzące od pewnego wytwórcy są prawie zawsze dobre. Można by to sprawdzić ponawiając doświadczenia opisanego typu.

Nie zamierzamy oczywiście zapisywać tych wszystkich wyników na dobro autokorelacji znakowej, możemy jednak stwierdzić, że korelogramy stały się punktem wyjścia do analizy technicznej.

Prace cytowane

- [1] M. G. Kendall, *The advanced theory of statistics*, London 1948, tom II, rozdz. 29 i 30, str. 363-441.
 [2] Z. A. Łomnicki, S. K. Zaremba, *Some applications for zero-one processes*, Journal of the Royal Statistical Society, series B, 17 (1955), str. 243-255.
 [3] H. Wold, *A study in the analysis of stationary time series*, Uppsala 1938.
 [4] — *Stationary time series*, Lucknow 1952.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 31. 12. 1956

П. МИКУЛЬСКИЙ, Я. ОДЕРФЕЛЬД, В. РУДСКИЙ (Варшава)

ЗНАКОВАЯ АВТОКОРРЕЛОГРАММА

РЕЗЮМЕ

Пусть x_1, x_2, \dots, x_n — последовательность действительных чисел, являющихся реализациями последовательности случайных переменных $\{X_i\}$. Обозначим через $E(X_i)$ ожидаемые значения, через y_i — разности $x_i - E(X_i)$.

Составляем последовательность знаков $\{\text{sign}(y_i y_{i+k})\}$. Пусть в этой последовательности имеется z_k знаков плюс. Выражение $\rho(k) = 2z_k/(n-k) - 1$ называем коэффициентом знаковой автокорреляции конечного ряда порядка k .

В этой статье доказано, что коэффициент $\rho(k)$ имеет подобные свойства, как классический коэффициент автокорреляции $r(k)$, а применение $\rho(k)$ значительно проще применения $r(k)$.

Если временной ряд составлен путем применения подвижной средней веса a_i ($i = 1, 2, \dots, m$) к последовательности независимых случайных величин ϵ_j с одинаковыми распределениями, то ожидаемое значение $\bar{\rho}(k) = E(\rho(k))$ и корреляционная функция в совокупности, обозначенная через $\bar{r}(k)$, связаны равенством $\bar{r}(k) = \sin[\bar{\rho}(k)\frac{1}{2}\pi]$.

Если временной ряд является так называемым авторегрессионным рядом, то $\bar{r}(k)$ и $\bar{\rho}(k)$ суть затухающие периодические функции с одинаковыми периодами и подобным затуханием амплитуд.

Эти результаты аналогичны результатам, полученным другим путем в работе [4]. Если временной ряд есть так называемый гармонический ряд, то $\bar{r}(k)$ является периодической функцией, а $\bar{\rho}(k)$ является с большим приближением периодической функцией того же периода.

В экспериментальной части построены классические и знаковые автокоррелограммы для временных рядов известной структуры и временных рядов, полученных из эксперимента, причем каждый раз получены хорошо согласованные результаты.

Наконец показано применение знаковой автокоррелограммы к диагнозу некоторых погрешностей в производстве папирос.

P. MIKULSKI, J. ODERFELD, W. RUDZKI (Warszawa)

SIGN CORRELOGRAM

SUMMARY

Let x_1, x_2, \dots, x_n be a sequence of real numbers which are realizations of a sequence of random variables $\{X_i\}$. Denote by $E(X_i)$ the expected values and by y_i the differences $x_i - E(X_i)$.

We form a sequence of signs $\{\text{sign}(y_i y_{i+k})\}$.

Suppose there are z_k „plus” signs in this sequence. The expression $\rho(k) = \frac{2z_k}{n-k} - 1$ is called the sign correlation coefficient of a finite series of order k .

In the present paper it is proved that the properties of $\rho(k)$ are similar to the properties of the classical, empirical correlation coefficient $r(k)$, and that the practical application of $\rho(k)$ is much simpler than that of $r(k)$.

If a time series is formed by applying a moving average with the weights a_i ($i = 1, 2, \dots, m$) to a sequence of independent random variables ε_j with identical distributions, then the relation between the expected value $\bar{\rho}(k) = E(\rho(k))$ and the correlation function in the population denoted by $\bar{r}(k)$ is $\bar{r}(k) = \sin[\bar{\rho}(k) \frac{1}{2}\pi]$.

If the time series is the so called autoregression series, then $\bar{r}(k)$ and $\bar{\rho}(k)$ are periodic damped functions with identical periods and a similar manner of damping the amplitudes.

These results are analogical to those obtained on different lines in [4].

If the time series is a so called harmonic series, then $\bar{r}(k)$ is a periodic function and $\bar{\rho}(k)$ is an approximately periodic function with the same period as $\bar{r}(k)$.

The experimental part contains the construction of classical and sign correlograms of time series with a known structure and of time series obtained from experiment, good agreement being obtained each time. Finally, the authors show the application of the sign correlogram to diagnose certain defects in the production of cigarettes.