

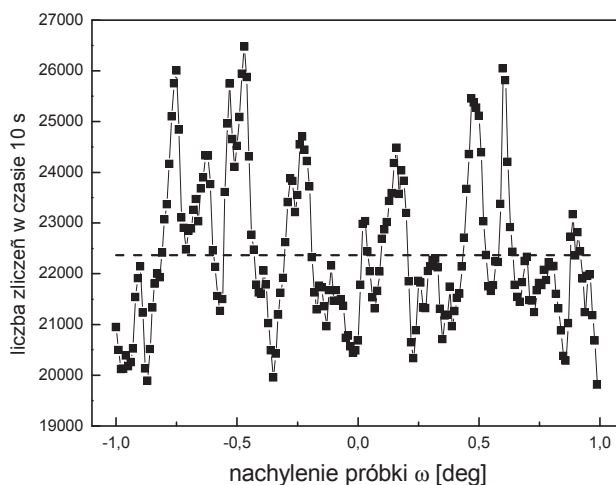
NIEPEWNOŚĆ POMIARU DLA CIĄGU OBSERWACJI SAMOSKORELOWANYCH

Andrzej Zięba
Akademia Górniczo-Hutnicza

1. Wstęp

Ciąg obserwacji $x_1, \dots, x_n$ jest zwykle rejestrowany w kolejności ich uzyskania. Kolejność ta jest bez znaczenia, jeżeli kolejne obserwacje x_i pozostają statystycznie niezależne od siebie, czyli *nieskorelowane*. Standardowe wzory na opracowanie serii obserwacji (np. [1]) są optymalne przy dodatkowych założeniach *równoważności* elementów próby i *normalności* ich rozkładu gęstości prawdopodobieństwa.

Jeżeli prawdopodobieństwo uzyskania wyniku x_i zależy od innych obserwacji, mamy do czynienia z *obserwacjami samoskorelowanymi* (*autocorrelated observations*). Rys. 1 przedstawia rzeczywiste dane samoskorelowane. Na oko wyglądają one inaczej niż dane nieskorelowane („biały szum”); są mniej poszarpane, zaś różnica $x_i - \bar{x}$ mniej często zmienia znak.



Rys. 1. Samoskorelowane dane z dyfraktometru rentgenowskiego: zależność natężenia refleksu 600 korundu (Al_2O_3) od kąta nachylenia próbki [2]. Linia przerywana oznacza średnią arytmetyczną $\bar{x}$ z pokazanych $n = 200$ wartości x_i

Samoskorelowane ciągi obserwacji uzyskiwane są w najróżniejszych eksperymentach. Najczęściej występują wtedy, gdy badany układ będący źródłem danych podlega losowym zakłóceniom i jednocześnie pamięta stan poprzedni. W efekcie mamy do czynienia z pewnym ciągłym procesem stochastycznym $x(t)$ (np. zależność ciśnienia

atmosferycznego od czasu), zaś dyskretny ciąg obserwacji x_i wynika z jego próbkowania w równych odstępach czasu.

Celem pracy jest przedstawienie właściwości statystycznych n -elementowej próby obserwacji samoskorelowanych i metod obliczenia dla niej niepewności typu A.

2. Matematyczne modele danych samoskorelowanych

Samoskorelowane ciągi liczb znane są w literaturze [3 ÷ 5] pod nazwą *szeregów czasowych*. W analogii do szeregów algebraicznych, szeregi czasowe są zdefiniowane za pomocą wzorów rekurencyjnych, wyrażających kolejne liczby x_i przy użyciu wartości poprzednich oraz ciągu liczb nieskorelowanych u_i . Dwa najprostsze przykłady to średnia ruchoma

$$\text{SMA: } x_i = \frac{u_i + u_{i-1} + \dots + u_{i-m}}{m} \quad (1)$$

oraz autoregresyjny szereg czasowy 1. rzędu

$$\text{AR(1): } x_i = a x_{i-1} + u_i \quad (2)$$

z parametrem $0 < a < 1$. (Symbole SMA oraz AR(1) pochodzą od określeń angielskich: *simple moving average* oraz *first-order autoregressive process*). Wzory (1), (2) i podobne umożliwiają łatwą generację liczb samoskorelowanych potrzebnych do obliczeń metodą Monte Carlo (MC).

W pracach autora [6 ÷ 8] sformułowany został pogląd, że opis teoretyczny rozpatrywanej kategorii danych samoskorelowanych jest możliwy przy wykorzystaniu prostszego formalizmu, polegającego na potraktowaniu ciągu obserwacji $x_1, \dots, x_n$ jako realizacji wielowymiarowej zmiennej losowej $X_1, \dots, X_n$. Przy podtrzymaniu założeń o normalności rozkładu i równoważności elementów próby losowej, statystyczne własności liczb x_i są w pełni scharakteryzowane przez wartość oczekiwaną μ oraz odchylenie standardowe σ . Zaś korelacje między X_i i X_j opisują elementy dyskretnej funkcji autokorelacji $\{\rho_k\}$, określone jako:

$$\rho_k = \frac{\text{cov}(X_i, X_{i+k})}{\sigma^2} \quad (3)$$

której wskaźnik $|k = i - j|$ nazywamy jest *odstępem*. Zatem, dla wyprowadzenia i zrozumienia przedstawionego w pracy formalizmu, wystarczy elementarny zbiór pojęć statystycznych zestawiony w Dodatku C *Przewodnika* [1].

Dla rzeczywistych danych samoskorelowanych występujące korelacje są prawie zawsze:

- (i) *dodatnie*, tzn. wszystkie $\rho_k \geq 0$, co wynika z faktu, że w różnych zjawiskach zwykle stan poprzedni jest podtrzymywany (pogoda dzisiejsza bywa podobna do wczorajszej),
- (ii) *o skończonym zasięgu*, tzn. elementy ρ_k maleją do zera ze wzrostem odstepu k (w perspektywie wielu dni korelacje parametrów meteorologicznych zanikają).

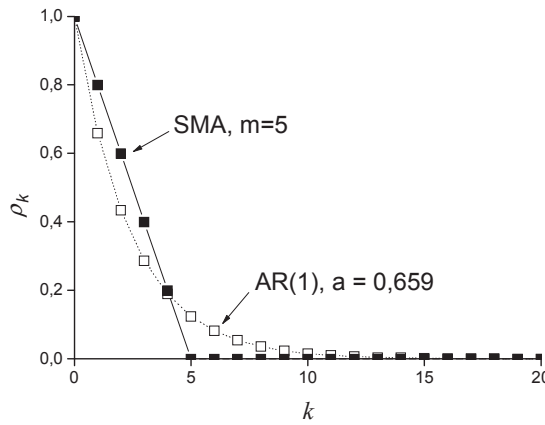
3. Funkcja autokorelacji: znana a priori i estymowana z danych

Funkcję $\{\rho_k\}$ jest znana *a priori* dla modelowych danych będących szeregami czasowymi generowanymi przez komputer. Dla szeregów czasowych (1) oraz (2) funkcję autokorelacji określają wzory

$$\text{SMA: } \begin{cases} \rho_k = 1 - k / m & \text{dla } k < m \\ \rho_k = 0 & \text{dla } k \geq m \end{cases} \quad (4)$$

$$\text{AR(1): } \rho_k = a^k \quad (5)$$

których wykresy przedstawiono na rys. 2.



Rys. 2. Funkcje autokorelacji dla dwu modeli szeregów czasowych

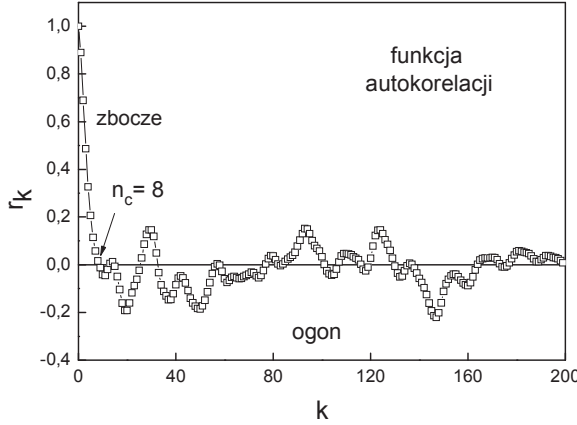
Istnieją sytuacje, gdy funkcja autokorelacji jest znana również dla danych uzyskanych z eksperymentu. Realizacją fizyczną procesu AR(1) jest np. przefiltrowanie szumu termicznego przez układ całkujący RC [9]. W pewnych przypadkach postać funkcji $\{\rho_k\}$ jest znana z dokładnych pomiarów wcześniejszych, dotyczy to np. danych meteorologicznych [10]. Wreszcie, funkcja autokorelacji jest znana, jeżeli dane samokorelowane powstały w wyniku wygładzania pierwotnych danych nieskorelowanych przy pomocy określonego algorytmu (np. obliczania średniej ruchomej (1)).

Najczęściej jednak funkcja autokorelacji musi być wyznaczona na podstawie posiadanych danych. Estymator funkcji autokorelacji $\{r_k\}$ zdefiniowany wzorem:

$$r_k = \frac{\sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

jest najczęściej omawiany w podręcznikach [3], [4] i [5], jak również zaimplementowany w programach do analizy danych. Rys. 3 przedstawia obliczoną przy pomocy wzoru (6) estymatę $\{r_k\}$ dla danych z rys. 1.

W wykresie estymaty można wyróżnić dwie, różne jakościowo części. Dla małych wartości odstępów k mamy opadające *zbocze*, w którym jest zawarta informacja o występujących korelacjach. Reszta to fluktuujący *ogon*. Jego takie czy inne kształty robią wrażenie quasi-ciągłych funkcji, jednakże jest to obraz szumu, tyle że samoskorelowanego. Odchylenie standardowe tych fluktuacji jest większe niż $n^{-1/2}$ [4], co tłumaczy dużą amplitudę fluktuacji w ogonie $\{r_k\}$.



Rys. 3. Estymata funkcji autokorelacji dla danych z rys. 1

4. Efektywna liczba obserwacji

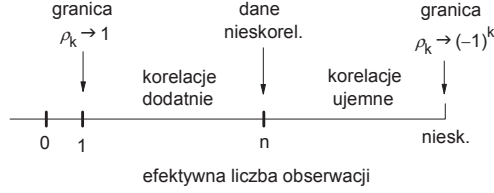
Dla opracowania rzeczywistych danych skorelowanych wprowadzono, w co najmniej dziesięciu pracach (wyszczególnionych w [8]) i niezależnie od siebie, parametr statystyczny nazwany *efektywną liczbę obserwacji* n_{eff} . Umożliwia on zwięzły zapis wzorów opisujących niepewność pojedynczego pomiaru i średniej (patrz pkt. 5), a ponadto pomaga w intuicyjnym zrozumieniu właściwości obserwacji skorelowanych. Przy znajomości funkcji autokorelacji $\{\rho_k\}$ wartość n_{eff} definiuje wzór:

$$n_{\text{eff}} = \frac{n}{1 + 2 \sum_{k=1}^{n-1} \frac{n-k}{n} \rho_k} \quad (7)$$

Rys. 4 przedstawia możliwe wartości efektywnej liczby obserwacji dla ustalonej liczbebnosci próby n . Ponieważ korelacje są na ogół dodatnie, to zwykle $n_{\text{eff}} < n$.

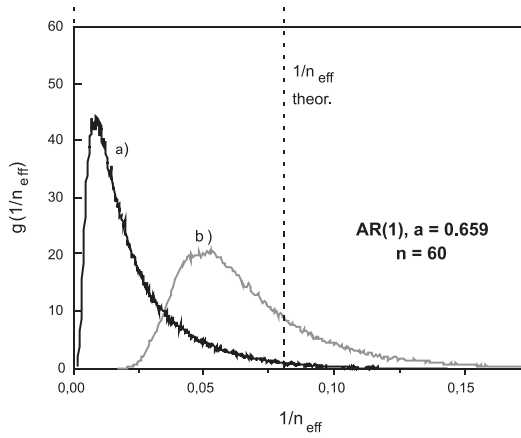
Jeżeli znamy tylko estymatę $\{r_k\}$ funkcji autokorelacji, powstaje zagadnienie, jak na jej podstawie estymować efektywną liczbę obserwacji? O ile n_{eff} obliczone na podstawie funkcji $\{\rho_k\}$ jest ustalonym parametrem, to z funkcji $\{r_k\}$ możemy otrzymać tylko estymator efektywnej liczby obserwacji $\hat{n}_{\text{eff}}$ będący, jak każdy estymator, zmienną losową o pewnej funkcji gęstości prawdopodobieństwa, wartości oczekiwanej i wariancji.

Narzucający się estymator efektywnej liczby obserwacji polega na użyciu wzoru (7), w którym funkcję $\{\rho_k\}$ zastępujemy przez $\{r_k\}$. Symulacja metodą MC pokazuje



Rys. 4. Przedstawienie właściwości n_{eff} na osi liczbowej. Granica liczb w pełni skorelowanych dodatnio oraz ujemnie otrzymuje się dla wartości ρ_k dążących do wartości równych, odpowiednio, 1 oraz $(-1)^k$

(rys. 5, krzywa a), że uzyskany w ten sposób estymator ma bardzo niekorzystne właściwości: wartość $1/\hat{n}_{\text{eff}}$ jest mocno zaniżona w stosunku do wartości teoretycznej, co prowadzi (wzór (11)) do zaniżenia oszacowania niepewności pomiaru.



Rys. 5. Funkcje gęstości prawdopodobieństwa dla różnych estymatorów odwrotności efektywnej liczby obserwacji: (a) sumowania wszystkich elementów r_k , b) wzór (7) z wartością n_c z metody FTZ. Symulacje MC wykonano dla modelu AR(1), $a = 0,659$, $n = 60$. Linia przerywana pokazuje wartość teoretyczną $1/n_{\text{eff}} = 1/12,33$

Przyczyną tego zjawiska jest wpływ ogona funkcji autokorelacji. Nieformalnym remedium jest arbitralne ograniczenie sumowania we wzorze (7) do tylko początkowych elementów estymaty $\{r_k\}$ [11]. Aby procedurę tą uczynić jednoznaczną, zaproponowano w [7] i [8] wzór:

$$\hat{n}_{\text{eff}} = \frac{n}{1 + 2 \sum_{k=1}^{n_c} \left(1 - \frac{k}{n}\right) r_k} \quad (8)$$

w którym graniczny odstęp n_c jest określony przez ostatni niezerowy element estymaty $\{r_k\}$ przed jej pierwszym przejściem przez zero (metoda FTZ – od ang. *first transit through zero*). Przykładowo, dla funkcji pokazanej na rys. 3 otrzymujemy $n_c = 8$.

Symulacyjne badanie MC pokazuje, że zastosowanie takiego podejścia daje radykalne zmniejszenie ujemnego obciążenia estymatora (rys. 5, krzywa b).

Dalsze zmniejszenie obciążenia estymatora uzyskuje się przy użyciu wzoru:

$$n_{\text{eff}}^+ = \frac{n - 2n_c - 1 + n_c(n_c + 1)/n}{1 + 2 \sum_{k=1}^{n_c} r_k} + 1 \quad (9)$$

wynikającego z użycia nowego estymatora efektywnej liczby obserwacji [12].

5. Średnia arytmetyczna jako wynik pomiaru powtarzanego

Średnia arytmetyczna jest najbardziej efektywnym przybliżeniem wartości oczekiwanej również dla przypadku, gdy elementy x_i próby losowej są wzajemnie skorelowane. Twierdzenie to można wywieść z metody najmniejszych kwadratów uogólnionej na przypadek skorelowanych danych wejściowych, przy czym dopasowaną funkcją jest funkcja stała. Brak wpływu korelacji na estymator wartości oczekiwanej można heurystycznie powiązać z faktem, że (i) średnia jest sumą zmiennych losowych X_i podzieloną przez n , oraz (ii) wartość oczekiwana sumy zmiennych losowych jest sumą wartości oczekiwanych, tak dla zmiennych niezależnych jak i skorelowanych.

6. Odchylenie standardowe pojedynczego pomiaru i średniej

Przedstawiona wyżej własność wartości oczekiwanej nie dotyczy wariancji. Wariancja sumy zmiennych losowych jest sumą wariancji poszczególnych składników sumy plus dodatkowe wyrazy zależne od współczynników korelacji między składnikami. Zatem prawidłowe wzory na estymatory wariancji i wariancji średniej muszą zależeć od elementów ρ_k funkcji autokorelacji.

W pracach [6] i [8] przedstawiono wyprowadzenie wzoru na nieobciążony estymator wariancji s_a^2 dla zmiennych samoskorelowanych¹. Poszukiwania literaturowe wykazały, że został on opublikowany bez wyprowadzenia przez Bayleya i Hammersleya [13], a następnie zapomniany. Wygodnie jest zapisać go w postaci wzoru:

$$s_a^2 = \frac{n_{\text{eff}}(n-1)}{n(n_{\text{eff}}-1)} s^2, \quad \text{gdzie} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (10)$$

jest „zwykłym” estymatorem wariancji. Ze wzrostem n i n_{eff} czynnik $\{n_{\text{eff}}(n-1)/[n(n_{\text{eff}}-1)]\}^{1/2}$ dąży do jedności, zatem zachodzi $s_a^2 \cong s^2$, ale dla próby o małej liczebności wzór (10) zapobiega zaniżeniu estymat wariancji i odchylenia standardowego $s_a = \sqrt{s_a^2}$.

Dla obserwacji niezależnych zmniejszenie odchylenia standardowego średniej $s(\bar{x})$ w stosunku do odchylenia standardowego s pomiaru o czynnik $n^{1/2}$ spowodowało

¹ Indeks a (od: *autocorrelated*) został wprowadzony, by odróżnić wprowadzone estymatory s_a^2 oraz $s_a^2(\bar{x})$ od dobrze znanych estymatorów s^2 i $s^2(\bar{x})$.

wane jest częściową kompensacją błędów o przeciwnych znakach. Dla zmiennych dodatnio skorelowanych liczba zmian znaku staje się mniejsza (rys. 1), zatem kompensacja błędu przypadkowego, wynikająca z częściowego znoszenia się przyczynków dodatnich i ujemnych, staje się mniej skuteczna. W rezultacie, w związku między niepewnościami pojedynczego pomiaru i średniej, zamiast n występuje mniejsza od niej efektywna liczba obserwacji, zatem:

$$s_a(\bar{x}) = \frac{s_a}{\sqrt{n_{\text{eff}}}} \quad (11)$$

W połączeniu z wzorem (10) niepewność pomiaru dla serii danych samoskorelowanych, utożsamianą z $s_a(\bar{x})$, zapisać można jako:

$$u(x) \equiv s_a(\bar{x}) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n_{\text{eff}} - 1)}} \quad (12)$$

Reasumując: w przypadku danych samokorelowanych możemy nadal wynik pomiaru utożsamiać ze średnią, natomiast do obliczania niepewności stosować należy wprowadzone wyżej wzory (10) ÷ (12).

Dokładność statystycznej oceny niepewności nie jest zbyt duża i zależy od parametru nazwanego liczbą stopni swobody $\nu = n - 1$. Dla danych samoskorelowanych można wprowadzić podobny parametr nazwany *efektywną (wypadkową) liczbą stopni swobody*. Do celów praktycznych wystarczy korzystać z wzoru przybliżonego [7, 8], zgodnie z którym:

$$\nu_{\text{eff}} \cong \frac{n}{1 + 2 \sum_{k=1}^{n-1} \rho_k^2} - 1 \quad (13)$$

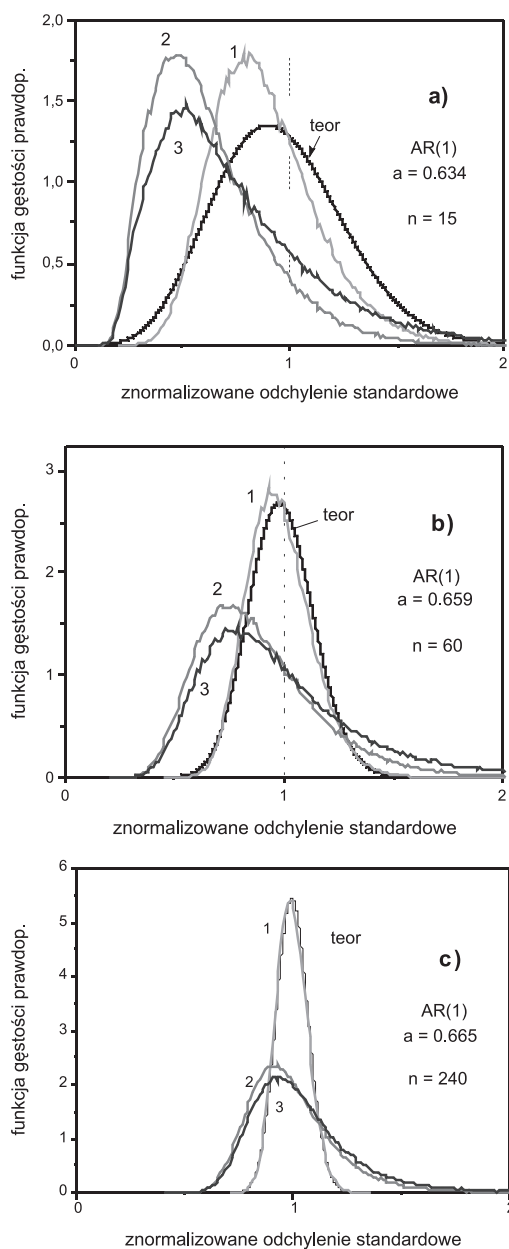
Wprowadzenie ν_{eff} umożliwia zapisanie wzoru na względną dyspersję estymatorów odchylenia standardowego w postaci:

$$\frac{u(s_a)}{s_a} = \frac{u(s_a(\bar{x}))}{s_a(\bar{x})} \cong \frac{1}{\sqrt{2\nu_{\text{eff}}}} \quad (14)$$

W granicy obserwacji niezależnych ($\rho_k = 0$ dla $k \geq 1$) otrzymuje się znany wzór $u(s)/s = [2(n - 1)]^{-1/2}$.

Przedstawiony formalizm analityczny (wzory (10) ÷ (14)) może być stosowany dla dowolnej liczebności próby i każdej funkcji autokorelacji w przypadku, gdy ta ostatnia jest znana. Gdy dysponujemy tylko estymatą $\{r_k\}$, to mamy dwa ograniczenia. Wyrażona wzorami (8) i (9) metoda estymowania efektywnej liczby obserwacji wymaga założenia, że elementy rzeczywistej funkcji autokorelacji są nieujemne i stają się praktycznie równe zeru dla $k < n$. Ponadto, rozrzut estymatora $\hat{n}_{\text{eff}}$ jest źródłem dodatkowego rozrzutu estymatorów odchylenia standardowego.

Ilościową informację o własnościach estymatorów odchylenia standardowego dla przypadku, gdy funkcja autokorelacji estymowana jest z danych, można uzyskać przy pomocy metody Monte Carlo. W pracach [7] i [12] przeprowadzono symulację



Rys. 6. Funkcje gęstości prawdopodobieństwa dla estymatorów znormalizowanego odchylenia standardowego dla modelu AR(1) i podanych trzech wielkości próby losowej. Krzywe 1, 2 oraz 3 pochodzą z symulacji MC i dotyczą, odpowiednio, s_a / σ , $s_a(\bar{x}) / \sigma(\bar{x})$ oraz $s_a^+(\bar{x}) / \sigma(\bar{x})$. Krzywe teoretyczne „teor” określa wzór (15) z $\nu \equiv \nu_{\text{eff}}$

dla modelu AR(1) i trzech licznosci próby losowej: $n = 15, 60$ i 240 , odpowiadających teoretycznym wartościom efektywnej liczby obserwacji $n_{\text{eff}} = 3,36, 12,33$ i $48,32$. Rys. 6 przedstawia uzyskane funkcje gęstości prawdopodobieństwa dla trzech estymatorów względnego odchylenia standardowego:

krzywa 1 – s_a / σ , gdzie do obliczenia s_a użyto kombinacji wzorów (10) i (8),
 krzywa 2 – $s_a(\bar{x}) / \sigma(\bar{x})$, gdzie $s_a(\bar{x})$ obliczano przy pomocy wzorów (11) i (8),
 krzywa 3 – $s_a^+(\bar{x}) / \sigma(\bar{x})$, gdzie $s_a^+(\bar{x})$ został obliczony z wzorów (11) i (9), czyli z wykorzystaniem nowego estymatora $\hat{n}_{\text{eff}}^+$ o obniżonym obciążeniu.

Dla obserwacji niezależnych znormalizowane odchylenie standardowe $z = s / \sigma$ podlega funkcji rozkładu prawdopodobieństwa:

$$g(z) = \frac{2^{1-\nu/2} \nu^{\nu/2}}{\Gamma(\nu/2)} z^{\nu-1} \exp(-\nu z^2 / 2) \quad (15)$$

wynikającej z rozkładu chi-kwadrat. Wzór (15) zastosowano dla obserwacji samoskorelowanych, przy czym za zmienną ν podstawiono teoretyczne wartości efektywnej (wypadkowej) liczby stopni swobody ν_{eff} dla badanych modeli (równe, odpowiednio, 5,4, 22,7 i 91,8). Obliczone krzywe teoretyczne pokazano na rys. 6.

Uzyskane metodą MC funkcje gęstości prawdopodobieństwa dla znormalizowanego odchylenia standardowego (krzywe 1 na rys. 6) wykazują dobrą zgodność z krzywymi teoretycznymi, zwłaszcza wtedy, gdy liczebność próby jest duża. Jest tak dlatego, że rozrzut estymatora $\hat{n}_{\text{eff}}$ ma niewielki wpływ na rozrzut s_a , gdyż $\hat{n}_{\text{eff}}$ wchodzi do wzoru (10) poprzez bliski jedności czynnik $\{n_{\text{eff}}(n-1)/[n(n_{\text{eff}}-1)]\}^{1/2}$. Dobra zgodność dowodzi, że efektywna (wypadkowa) liczba stopni swobody ν_{eff} jest parametrem dobrze opisującym rozrzut estymatorów wariancji i odchylenia standardowego.

Wpływ losowego charakteru estymatorów efektywnej liczby obserwacji $\hat{n}_{\text{eff}}$ jest wyraźny dla odchylenia standardowego średniej (czyli niepewności pomiaru). Efektem jest zaniżenie wartości oczekiwanej tego estymatora (szczególnie dla małej liczebności próby, rys. 6a, b) i powiększony rozrzut. Jednak, biorąc pod uwagę fakt, że niepewność pomiaru znamy w ogólności z niewielką dokładnością, przedstawiony formalizm może być stosowany w praktyce.

7. Konkluzje

Rezultatem pracy jest przedstawienie formalizmu, który rozszerza zwykłą metodę obliczania niepewności standardowej typu A na przypadek danych samoskorelowanych. Formalizm ten opiera się w na istniejących rezultatach teoretycznych, w większości mało znanych, uzupełnionych przez nowe koncepcje i przetestowaniu całości formalizmu przy użyciu modelowania Monte Carlo.

Bezpośrednim przedłużeniem pracy, doprowadzonej do zagadnienia niepewności standardowej, będzie zbadanie zagadnienia szacowania niepewności rozszerzonej. Inne możliwości zastosowań przedstawionej metodologii to testy statystyczne dla danych samoskorelowanych, stacjonarne procesy stochastyczne z nałożonym trendem czy wreszcie niestacjonarne procesy stochastyczne.

Literatura

- [1] *Guide to the Expression of Uncertainty in Measurement*, ISO 1995. Tłumaczenie polskie: *Wyrażanie Niepewności pomiaru*. Przewodnik. Główny Urząd Miar, 1999.
- [2] A. Zięba: *Niepewność pomiaru obserwacji skorelowanych*. VII Sympozjum nt. Niepewności pomiarów i XXI Seminarium Sekcji Podstaw Metrologii KMiAN PAN, Świnoujście, 11 – 15 luty 2008.
- [3] G. E. P. Box, G. M. Jenkins: *Analiza szeregów czasowych. Prognozowanie i sterowanie*. PWN, Warszawa 1983 (Nowsze wydanie z rozszerzonym zespołem autorów: G. E. P. Box, G. M. Jenkins, G. C. Reinse. *Time Series Analysis: Forecasting and Control*. 1994, Prentice Hall 1994).
- [4] M. B. Priestley: *Spectral Analysis and Time Series*. Elsevier, 1981.
- [5] P. J. Brockwell, R. A. Davis: *Time series: theory and methods* 2nd ed. Springer, New York, 1991.
- [6] A. Zięba: *Niepewność wartości średniej serii obserwacji skorelowanych*. Podstawowe Problemy Metrologii PPM'08, Sucha Beskidzka 11 – 14 maj 2008. Materiały konferencji wydane przez Komisję Metrologii Oddziału PAN w Katowicach.
- [7] A. Zięba, P. Ramza: *Niepewność wartości średniej serii obserwacji skorelowanych (II)*. Podstawowe Problemy Metrologii PPM'09, Sucha Beskidzka 10 – 13 maj 2009. Materiały konferencji wydane przez Komisję Metrologii Oddziału PAN w Katowicach.
- [8] A. Zięba: *Effective number of observations and unbiased estimators of variance for autocorrelated data – an overview*. Metrology and Measurement Systems, 2010, vol. 17, 3-16.
- [9] T. J. Witt: *Using the autocorrelation function to characterize time series of voltage measurements*. Metrologia, 2007, vol. 44, pp. 201-209.
- [10] C. E. Leith: *The standard error of time-averaged estimates of climatic means*. J. Appl. Meteorol., 1973, vol. 12, 1066-1069.
- [11] M. Dorozhovets, Z. L. Warsza: *Udoskonalenie metod wyznaczania niepewności wyniku pomiaru w praktyce*. Przegląd Elektrotechniczny, 2007, vol. 83, 1-13.
- [12] A. Zięba, P. Ramza: manuskrypt wysłany do Statistics and Computation.
- [13] G. V. Bayley, G. M. Hammersley: *The "Effective" Number of Independent Observations in an Autocorrelated Time-Series*. J. R. Stat. Soc. Suppl., 1946, 8, 184-197.