

УДК 621.317

ЗАСТОСУВАННЯ МЕТОДУ СЕРІЙ ДЛЯ ДОСЛІДЖЕННЯ ВЗАЄМНОЇ КОРЕЛЯЦІЇ СПОСТЕРЕЖЕНЬ

© Дорожовець Михайло, Никипанчук Олена, 2016

Національний університет “Львівська політехніка”,

кафедра інформаційно-вимірювальних технологій, вул. С. Бандери 12, 79013, Львів, Україна

Запропоновано метод обчислення автокореляції, а також ефективної кількості спостережень. Для генерування корельованих спостережень використовується метод рухомого середнього. За допомогою методу серій можливе спрощене обчислення ефективної кількості спостережень для визначення стандартної непевності середнього значення корельованих спостережень.

Ключові слова: спостереження, кореляція, непевність, метод серій, ефективне число.

Предложен метод вычисления автокорреляции, а также эффективного числа наблюдений. Для генерирования коррелированных наблюдений используется метод подвижного среднего. С помощью метода серий возможно упрощенное вычисление эффективного числа наблюдений для определения стандартной неопределенности среднего значения коррелированных наблюдений.

Ключевые слова: наблюдение, корреляция, неопределенность, метод серий, эффективное число.

When processing the results of measurements big role important presence of correlation values. To find the standard uncertainty need to know the effective number of uncorrelated observations. No correlation can consider could lead to incorrect evaluation of the standard uncertainty of the mean.

Not always known autocorrelation function monitoring, and evaluation of the autocorrelation function on observations characterized by low accuracy, which can lead to incorrect finding effective number.

There are indirect methods of evaluating the impact assessment on observations correlation standard deviation. This method is recorded sample of N divided into k sub-samples (groups) up to n samples each ($N = n \cdot k$). Each subsample are partial mean and variance estimation, and find the settings for the entire sample. Then compare the ratio of the variance between groups and within the group. Using the F distribution at a significance level α determined whether the observations are correlated or not. These methods are quite complex and require significant additional computing.

The purpose of research is to study simple method of testing autocorrelation and consideration in calculating the N_{eff} . The proposed method is based on calculating the number of series. Series is a sequence of observed values equal before which or after which the values observed are another category or no supervision at all. Set the number of series or observation results are correlated or not. To determine whether correlated observations required to determine the median of the sample and calculate the number of deviations from the median values.

Research performed by the Monte Carlo. For research use two types of observations: first – with uncorrelated observations, the second – generated correlated observations, including the method of moving average. To find the index of correlation function used exponential autocorrelation function.

An effective dependence theoretical number and effective number determined by the method episodes from different bias moving average on a constant number of observations.

Based on these studies show that increasing the number of observations ($N > 50$) to simplify the calculation of the possible number of effective using the method of series. At least 50 the number of observations can be effective calculating numbers with a small bias moving average. To investigate the cross-correlation of observations of the method is appropriate series and simplifies the calculation of the standard uncertainty.

Key words: monitoring, correlation, uncertainty, series method, the effective number.

Вступ. Для опрацювання результатів вимірювання велику роль має значення наявності корельованості значень. Вперше дослідив вплив корельованих спостережень ще в 1935 р. Дж. Бартельс [1], пізніше на

деякий час ці дослідження були “забуті” [2]. Тому сьогодні ця проблема ще актуальна.

Коли спостереження некорельовані, стандартна непевність середнього значення становить:

$$u_A(\bar{x}) = \frac{S_x}{\sqrt{N}}. \quad (1)$$

У випадку корельованих спостережень вона змінює своє значення [2–6]:

$$u_A(\bar{x}) = \frac{S_x}{\sqrt{N}} \sqrt{\frac{N-1}{N_{eff}-1}}. \quad (2)$$

Тому для знаходження стандартної непевності потрібно знати ефективну кількість некорельованих спостережень [3–6].

Теоретично ефективна кількість некорельованих спостережень визначається значенням коефіцієнтів r_i кореляційної функції [3–6]:

$$N_{eff} = \frac{N}{1 + 2 \sum_{i=1}^{N-1} (1 - \frac{i}{N}) r_i} = \frac{N}{1 + D_{r,N}}. \quad (3)$$

Невраховування кореляції може призвести до неправильної оцінки стандартної непевності середнього значення.

Не завжди відома функція автокореляції спостереження, а оцінювання впливу автокореляційної функції на результати спостережень характеризується невеликою точністю, що може призвести до неправильного знаходження ефективного числа [7].

Існують опосередковані методи оцінювання впливу корельованості спостережень на оцінювання стандартного відхилення.

Суть цього методу полягає в тому, що зареєстровану вибірку обсягом N ділять на k підвбірок (групи) обсягом по n вибірок кожна ($N=n \cdot k$) [6]. В кожній підвбірці знаходять часткове середнє значення та оцінку дисперсії, а також визначають ці параметри для усієї вибірки. Далі порівнюються відношення дисперсії між групами і в середині групи. Використовуючи розподіл F на рівні значущості α , визначають, чи спостереження корельовані, чи ні. Якщо спостереження корельовані, то на основі рівня відношення дисперсії посередньо можна обчислити N_{eff} і врахувати його під час оцінювання стандартної непевності.

Цей метод є достатньо складним і потребує значних додаткових обчислень.

Метою досліджень є опрацювання простого методу тестування автокореляції та врахування під час обчислення N_{eff} .

Теоретичні основи. Запропонований метод ґрунтується на підрахуванні кількості серій [8].

Серією називається послідовність однакових спостережених значень, перед якими або після яких розташовані спостережені значення іншої категорії або спостереження відсутні ззагалі [8].

Кількість серій дає змогу встановити, чи є результати спостереження корельованими, чи ні.

Для некорельованих спостережень середнє значення кількості серій спостережень двох видів: одні більші від медіани, а інші менші від неї, дорівнює [8]:

$$m = \frac{N}{2} + 1, \quad (4)$$

а дисперсія –

$$s^2 = \frac{N(N-2)}{4(N-1)}, \quad (5)$$

де N – кількість спостережень.

Щоб визначити, чи спостереження корельовані, потрібно визначити медіану вибірки та підрахувати кількість відхилень значень від медіани.

Кількість серій для корельованих спостережень обчислюється згідно з виразом [8]:

$$m_k = \frac{2N_1N_2}{N} + 1, \quad (6)$$

а дисперсія –

$$s_k^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}, \quad (7)$$

де N_1 – значення, яке відхиляється від медіани зі знаком «+», N_2 – значення зі знаком «-».

Методика досліджень. Для дослідження використано два види спостережень: перший – з некорельованими спостереженнями, другий – з корельованими спостереженнями, згенерованими методом рухомого середнього. Дослідження виконано методом Монте-Карло з кількістю повторень експериментів M .

Некорельовані спостереження з нормальним законом розподілу згенеровано з використанням стандартних функцій MathCad.

Для отримання скорельованих спостережень спочатку згенеровано M вибірок довжиною $L = N + k$ кожна спостережень $x_{j,l}$ з нормальним законом розподілу, де $j=1 \dots M$, N – кількість спостережень ($i=1 \dots N$), яка змінювалася від 25 до 100, k – кількість усереднених некорельованих спостережень, що змінюється від 2 до 10.

Для спрощення в обох випадках приймали, що некорельовані спостереження $x_{i,j}$ мають нульове математичне сподівання ($m_x = 0$) та одинкове стандартне відхилення ($s_x = 1$).

Далі для заданої кількості спостережень N ($i = 1 \dots N$) вибірки з корельованими спостереженнями згенеровано за допомогою методу рухомого середнього згідно з виразом [8, 9]:

$$y_{i,j} = \frac{1}{k} \sum_{i1=1}^k x_{i+i1,j} \quad (8)$$

Теоретична функція автокореляції для вибірки, згенерованої за допомогою методу рухомого середнього (8), описується залежністю [7]:

$$r_i = \left(1 - \frac{i}{k}\right), \quad i = 0 \dots k \quad (9)$$

Подальші обчислення проводили однаково для двох вибірок: з некорельованими і корельованими спостереженнями.

Спочатку визначали медіани обох вибірок [10]:

$$X_{\text{мед},j}(N) = \begin{cases} X_{((N+1)/2),j}, & N - \text{непарне}; \\ \frac{X_{(N/2),j} + X_{\left(\frac{N}{2}+1\right),j}}{2}, & N - \text{парне}. \end{cases} \quad (10)$$

$$Y_{\text{мед},j}(N) = \begin{cases} Y_{((N+1)/2),j}, & N - \text{непарне}; \\ \frac{Y_{(N/2),j} + Y_{\left(\frac{N}{2}+1\right),j}}{2}, & N - \text{парне}. \end{cases} \quad (10)$$

Далі для кожної з M вибірок, порівнюючи, фіксували спостереження, які були більшими чи меншими від медіани:

$$a_{i,j} = \begin{cases} 1 - \text{якщо } -X_{i,j} \geq X_{\text{мед},j}; \\ 0 - \text{якщо } -X_{i,j} < X_{\text{мед},j}. \end{cases}$$

$$b_{i,j} = \begin{cases} 1 - \text{якщо } -Y_{i,j} \geq Y_{\text{мед},j}; \\ 0 - \text{якщо } -Y_{i,j} < Y_{\text{мед},j}. \end{cases} \quad (11)$$

Проаналізувавши значення $a_{i,j}$ та $b_{i,j}$, обчислили кількості серій відхилень спостережень від медіани:

$$Ks_j = \text{Numberserii}(a_{i,j}) - \text{для некорельованих спостережень}; \quad (12a)$$

$$Ksk_j = \text{Numberserii}(b_{i,j}) - \text{для корельованих спостережень}. \quad (12b)$$

Усреднюючи обчислені кількості серій (12), знаходили їхні середні значення – очікувані значення:

$$KS = \frac{1}{M} \sum_{j=1}^M Ks_j - KSk = \frac{1}{M} \sum_{j=1}^M Ksk_j \quad (13)$$

відповідно для некорельованих і корельованих спостережень.

За відомою нормованою функцією автокореляції (9) обчислювали показник впливу кореляції на дисперсію середнього значення y (3):

$$D_{r,N} = 2 \sum_{i=1}^{N-1} \left(1 - \frac{i}{N}\right) \cdot r_i. \quad (14)$$

Для знаходження показника кореляційної функції використовується трикутна функція автокореляції.

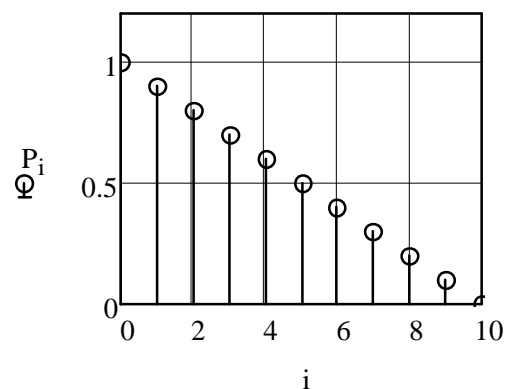


Рис. 1. Трикутна функція автокореляції

Fig. 1. The triangular autocorrelation function

На підставі (14) визначали теоретично ефективну кількість спостережень (3):

$$N_{\text{eff}} = \frac{N}{1 + D_{r,N}}. \quad (15)$$

Відповідно на підставі очікуваних значень кількості серій (13) за методом серій обчислювали ефективну кількість спостережень для корельованих спостережень:

$$N_{\text{eff},Ks} = \frac{N \cdot KSk}{KS}. \quad (16)$$

Результати дослідження. Залежність теоретично ефективної кількості (15) і ефективної кількості, визначеної за допомогою методу серій (16) для різної кількості спостережень (N) та кількості (k) усереднюваних спостережень у рухомій середній (8) наведено на рис. 2.

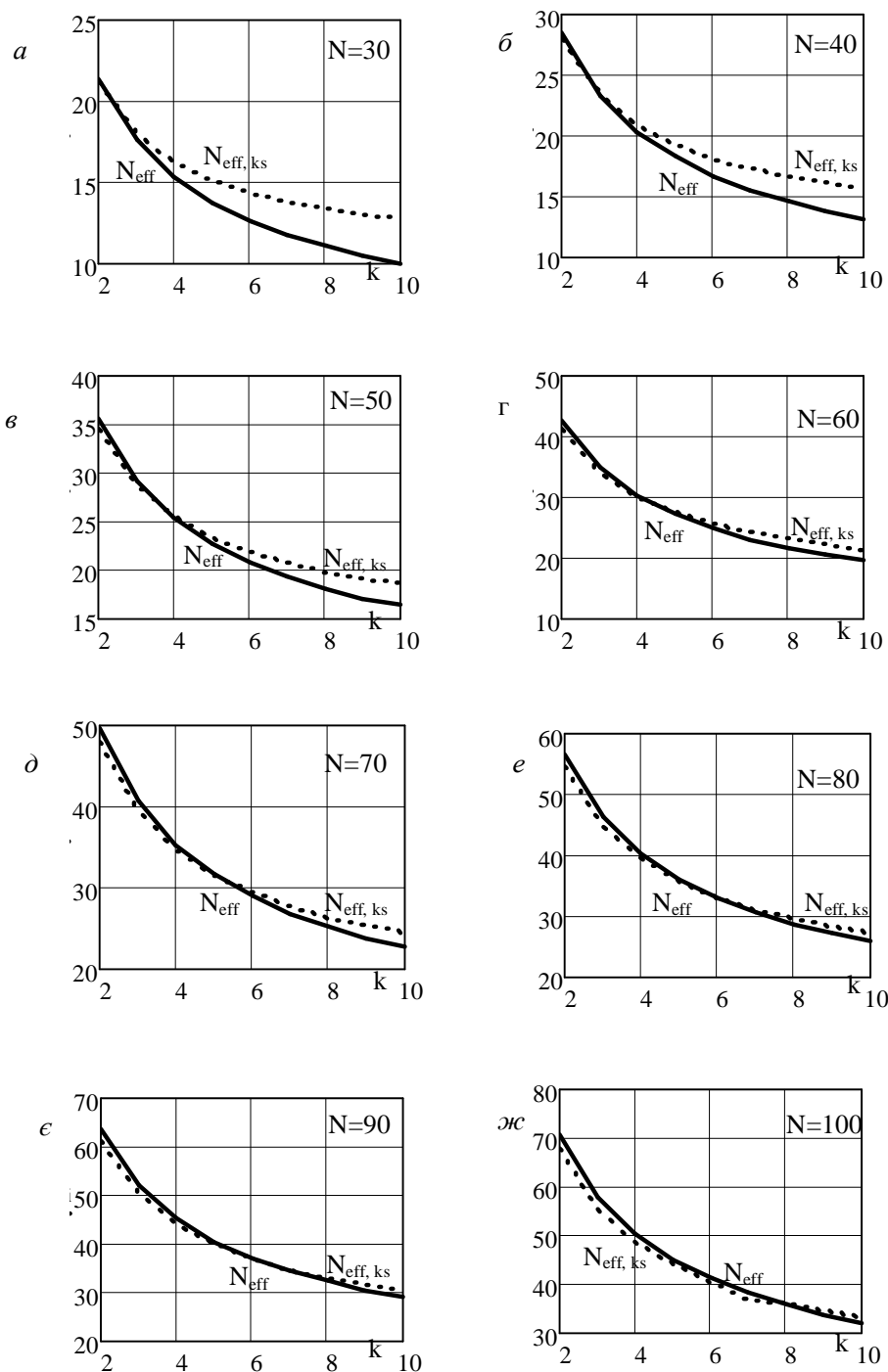


Рис. 2. Залежність теоретично ефективної кількості й ефективної кількості, визначеної за допомогою методу серій, від різних зміщень рухомого середнього за сталої кількості спостережень

Fig. 2. Dependence of the theoretical number of efficient and effective number determined by the method episodes from different bias moving average on a constant number of observations

Аналіз результатів. З отриманих і поданих на рис. 2 залежностей бачимо, що коли кількість спостережень перевищує 50, значення теоретично ефективної кількості й ефективної кількості, визначеної за допомо-

гою методу серій, наближені один до одного. Якщо кількість спостережень менша за 50, значення ефективної кількості наближені тільки за невеликого зміщення рухомого числа (до $k \approx N/10$).

Висновок. На основі цих досліджень показано, що у разі збільшення кількості спостережень ($N \approx \geq 50$) для спрощення можливе обчислення ефективного числа за допомогою методу серій.

Якщо кількість спостережень менша, ніж 50, можливе обчислення ефективного числа за невеликого зміщення рухомого середнього.

1. Bartels J. Zur Morphologie geophysikalischer Ziefunktionen. Sitz-Ber. Preuß. Akad. Wiss(1935) 30, P. 502–522. 2. Zięba A., Ramza P. Niepewność wartości średniej serii obserwacji skorelowanych // Materiały konferencji Podstawowe Problemy Metrologii PPM'09 (2009), Sucha Beskidzka, 11–14 maja 2009. – S. 80–84. 3. Dorozhovets M., Warsza Z. L. Wyznaczanie niepewności typu A pomiarów o skorelowanych rezultatach obserwacji. Pomiar, Automatyka, Kontrola(2007), No. 2. – S. 20–24. 4. Zhang N. F. Calculation of the uncertainty of the mean of

autocorrelated measurements // Metrology (2006) 4. – P. 276–281. Warsza Z. L., Dorozhovets M. Uncertainty type A evaluation of autocorrelated measurement observations. Bulletin WAT(2008) VOL. LVII, No. 2. – P. 141–152. 6. Dorozhovets M. Metoda pośredniego testowania wzajemnego skorelowania obserwacji losowych. VI Kongres metrologii. Mechanik, NR 7/2013. 7. Бокс Дж. Анализ временных рядов прогноз и управление (часть 1)// Дж. Бокс, Г. Дженкинс. – М., 1974. – 405 с. 8. Бендат Дж. Измерение и анализ случайных процессов // Дж. Бендат, А. Пирсол; пер. с англ. Г. В. Матушевского, В. Е. Привальского. – М.: Мир, 1971. – 408 с. 9. Большев Л. Таблицы математической статистики // Л. Н. Большев, Н. В. Смирнов. – М.: Наука, 1983. – 416 с. 10. Айвазян С. Прикладная статистика. Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М.: Финансы и статистика, 1983. – 472 с.