

ДОСЛІДЖЕННЯ МЕТОДУ СЕРІЙ ДЛЯ ОЦІНКИ ВЗАЄМНОЇ КОРЕЛЯЦІЇ СПОСТЕРЕЖЕНЬ ГЕНЕРОВАНИХ ЗА ДОПОМОГОЮ МЕТОДУ АВТОРЕГРЕСІЇ

© О. Никипанчук, М. Дорожовець, 2017

Національний університет „Львівська політехніка”, Львів, Україна

Наявність корельованості значень відіграє велику роль при опрацюванні результатів вимірювання. Ця проблема є актуальною, оскільки, дослідження впливу корельованих спостережень вперше були представлені ще в 1935 році Дж. Бартельс [1], але пізніше на деякий час ці дослідження були „забуті”.

Стандартна непевність середнього значення у випадку коли спостереження не є корельовані становить [2]:

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

У випадку коли спостереження корельовані стандартна непевність середнього значення змінюється пропорційно не від кількості спостережень, а від так званого ефективного числа не скорельованих спостережень.

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

Вперше термін „ефективне число незалежних спостережень” було запропоновано Бейлі та Хамерслі [3].

Ефективне число визначається за виразом:

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

Для оцінювання впливу корельованості спостережень на оцінювання стандартного відхилення існують опосередковані методи. Ці методи вимагають значних додаткових обчислень і є достатньо складними [4].

Методика досліджень

Для дослідження використано два види спостережень: 1-ий – з не корельованими спостереженнями; 2-ий – зкорельовані спостереження генеровані авторегресією. Дослідження виконані методом Монте Карло з кількістю експериментів $M = 10^5$. Згенеровано вибірку з нормальним законом розподілу для не корельованих спостережень використовуючи стандартні функції MathCad.

Для отримання скорельованих спостережень спочатку згенеровано M вибірок довжиною N кожна спостережень $x_{j,i}$ з нормальним законом розподілу де $j=1\dots M$, $i=1\dots N$, де N – кількість спостережень, яке змінювалося від 20 до 100. Для спрощення приймалося що параметр положення рівний $m_x=0$ та стандартне відхилення $\sigma_x=1$.

Для заданої кількості спостережень N вибірки з корельованими спостереженнями згенеровано за допомогою методу авторегресії [5,6]:

$$y_{j,i} = a \cdot x_{j,i} + b \cdot y_{j-1,i} \quad (4)$$

де $y_{0,i} = 0$, $a = \sigma_x \sqrt{1-b^2}$, $b = \rho_i$.

Теоретична функція автокореляції для вибірки генерованої за допомогою авторегресії описується залежністю

$$\rho_i = e^{-\alpha \cdot i}, \text{ де } i = 0 \dots \infty \quad (5)$$

для дослідження прийнято $\alpha=0,7$.

Запропонований метод ґрунтується на підрахуванні кількості серій [7].

Наступні обчислення проводили однаково для двох вибірок: з не корельованими і корельованими спостереженнями. Спочатку визначали медіани обох вибірок:

$$X(Y)_{med} = \begin{cases} X(Y)_{((N+1)/2),j}, & N - \text{непарне}; \\ X(Y)_{\frac{N}{2},j}, & N - \text{парне}. \end{cases} \quad (6)$$

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

$$a_{i,j} = \begin{cases} 1 - \text{якщо } -X_{i,j} \geq X_{\text{мед},j} ; \\ 0 - \text{якщо } -X_{i,j} < X_{\text{мед},j} . \end{cases} \quad b_{i,j} = \begin{cases} 1 - \text{якщо } -Y_{i,j} \geq Y_{\text{мед},j} ; \\ 0 - \text{якщо } -Y_{i,j} < Y_{\text{мед},j} . \end{cases} \quad (7)$$

Шляхом аналізу значень $a_{i,j}$ та $b_{i,j}$ обчислили кількості серій відхилень спостережень від медіани:

$$\mu_{aj} = \sum_{i=1}^N a_{i,j} - \text{для некорельованих спостережень, (8,а)} \quad \mu_{bj} = \sum_{i=1}^N b_{i,j} - \text{для корельованих спостережень, (8,б)}$$

Усреднюючи обчислені кількості серій (8) знаходили їх середні значення – очікувані значення:

$$\bar{\mu}_a = \frac{1}{M} \sum_{j=1}^M \mu_{aj} \quad \bar{\mu}_b = \frac{1}{M} \sum_{j=1}^M \mu_{bj} \quad (9)$$

відповідно для некорельованих і корельованих спостережень.

За відомою нормованою функцією автокореляції (5) обчислювали показник впливу кореляції на дисперсію середнього значення у (3):

$$D_{\rho,N} = 2 \sum_{i=1}^{N-1} \left(1 - \frac{i}{N}\right) \cdot \rho_i \quad (10)$$

На підставі (10) визначали теоретичне ефективне число спостережень (3):

$$N_{\text{eff}} = \frac{N}{1 + D_{\rho,N}} \quad (11)$$

Відповідно на підставі очікуваних значень кількості серій (9) згідно методу серій обчислювали ефективне число спостережень для корельованих спостережень:

$$N_{\text{eff},\rho} = \left(\frac{\mu_b}{\mu_a}\right)^2 \cdot N \quad (12)$$

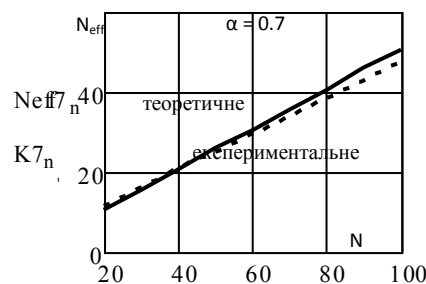


Рис.1. Залежність теоретичної ефективної кількості спостережень і ефективної кількості спостережень генерованих за допомогою методу авторегресії

Висновок.

Значення теоретичної кількості спостережень і кількості спостережень, які визначенні за допомогою методу серій і були генеровані за методом авторегресії є наближені один до одного при будь якому значенні кількості спостережень. Тому для спрощення доцільно використовувати метод серій для визначення ефективної кількості спостережень при оцінюванні стандартної непевності середнього значення.

1. Guide to the Expression of Uncertainty in Measurement, ISO 1992, revised and corrected 1995. 2. Bartels J. Zur Morphologie geophysikalischer Zeitfunktionen. Sitz-Ber. Preuß. Akad. Wiss(1935) 30, p. 502÷522. 3. G.V. Bayley, G.M. Hammersley, —“The —effective” number of independent observations in an autocorrelated time-series”, J. Roy. Stat. Soc. Suppl. 8, 1946, pp.184-197. 4. M. Dorozhovets. Metoda posredniego testovania wzajemnego skorelowania obserwacji losowych VI CONGRESS METROLOGY. MECHANIC No 7/2013. 5. Л. Большев. Таблицы математической статистики.//Л.Н. Большев, Н.В.Смирнов. – Москва: „Наука”, 1983. – 416 с. 6. R. Hanus, A. Szlachta, A. Kowalchuk. Generowanie sygnałów stochastycznych o zadanych charakterystykach statystycznych w badaniach metod estymacji czasu opóźnienia vol.57, PAK, nr 2/2011. 6. J Box, George E. P. Time series analysis : forecasting and control / George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel. —4th ed., John Wiley & Sons Hoboken, New Jersey 2008