МОДЕЛЮВАННЯ ЗАЛЕЖНИХ ГАУСІВСЬКИХ ВИПАДКОВИХ ВЕЛИЧИН НА ОСНОВІ ДЕКОМПОЗИЦІЇ КОРЕЛЯЦІЙНОЇ МАТРИЦІ ТА ПЕРЕТВОРЕННЯ ДЖОНСОНА

УДК 004:519.2:519.6

ПРИХОДЬКО Сергій Борисович
д.т.н., доцент, завідувач кафедри програмного забезпечення автоматизованих систем
Національного університету кораблебудування ім. адм. Макарова.
Наукові інтереси: математичне моделювання випадкових процесів в інформаційних технологіях.
e-mail: sergiy.prykhodko@nuos.edu.ua

ВСТУП

Для моделювання стохастичних систем, процесів та полів, як правило, виникає потреба в значеннях залежних гаусівських випадкових величин [1-3]. Зараз для моделювання залежних гаусівських випадкових величин відомі різні методи, які можна поділити на дві групи: методи, що базуються на виключеннях [4-6], та методи, що використовують різноманітні перетворення [7-9]. Методи на основі декомпозиції кореляційної матриці відносяться до другої групи, На відміну від методів першої групи вони не використовують значень відповідних багатомірних функцій розподілу чи щільності ймовірності, але потребують значень незалежних гаусівських випадкових величин. В свою чергу для моделювання незалежних гаусівських випадкових величин існує багато методів, які базуються на перетворенні випадкових чисел з рівномірним розподілом у такі, що мають розподіл Гауса [4-6, 10, 11]. Частка з цих методів базується на центральній граничній теоремі і потребує для генерації одного значення гаусівської випадкової величини від 12 значень випадкових чисел з рівномірним розподілом [2]. Інші – використовують спеціальні перетворення (наприклад, Бокса-Мюллера) і потребують на створення певної кількості значень гаусівської випадкової величини приблизно в 1,25 рази більше значень випадкових чисел з рівномірним розподілом [10]. Все це приводить до проблеми додаткових витрат ресурсів комп'ютера на моделювання. Крім того виникає проблема зменшення фактичної кількості значень псевдовипадкових чисел з рівномірним розподілом в межах періоду генератора, яку можна використовувати до їх повторення [11]. В [11] запропоновано метод моделювання незалежних гаусівських випадкових величин, який дозволяє зменшити витрати ресурсів комп'ютера на відповідне моделювання та максимально використовувати наявні можливості генераторів псевдовипадкових чисел за рахунок того, що для створення одного значення гаусівської випадкової величини необхідне тільки одне значення випадкової величини з рівномірним розподілом. Для цього в [11] застосовано нормалізуюче перетворення Джонсона із сім'ї S_B.

Ціль даної роботи полягає у удосконаленні метода моделювання залежних гаусівських випадкових величин на основі декомпозиції кореляційної матриці за рахунок використання нормалізуючого перетворення Джонсона із сім'ї S_B.

ВИКЛАДЕННЯ ОСНОВНОГО МАТЕРІАЛУ

В загальному випадку для моделювання залежних (корельованих) гаусівських випадкових величин з кореляційною матрицею **C** потрібно визначити матрицю U, таку що [9]

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{C} . \qquad (1)$$

За матрицею **U** та некорельованими випадковими числами z з розподілом Гауса можна отримати значення $\mathbf{z_C}$ залежних (корельованих) гаусівських випадкових величин

$$\mathbf{z_C} = \mathbf{zU} \,. \tag{2}$$

Тут **z** – це вектор-рядок з кількістю елементів, яка дорівнює числу залежних гаусівських випадкових величин. Елементи вектора z – це значення незалежних гаусівських випадкових величин.

Значення незалежної гаусівської випадкової величини з математичним сподіванням нуль і дисперсією одиниця визначаємо за перетворенням Джонсона із сім'ї S_B за формулою [11]

$$z = 0{,}3522056 \ln\left[\tilde{x}/(1-\tilde{x})\right], \tag{3}$$

де

$$\tilde{x} = \left(\sin\frac{\pi u}{2} + 1{,}0017416\right) \Big/ 2{,}0034832 \,.$$

Тут u – випадкова величина з рівномірним розподілом, $u \in [-1, 1]$.

Одним із загальних методів визначення матриці **U** з (1) є декомпозиція Холеского

$$\mathbf{U} = \mathbf{L}^{\mathrm{T}} \,,$$

де **L** – нижня трикутня матриця розкладання $\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, елементи якої визначаються як

$$L_{jj} = \sqrt{C_{jj} - \sum_{k=1}^{j-1} L_{jk}^2} \,;$$

$$L_{ij} = \frac{1}{L_{jj}}\left(C_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk}\right) \text{ для } i > j \,.$$

Розглянемо приклад моделювання трьох залежних гаусівських випадкових величин з кореляційною матрицею **C**

$$\mathbf{C} = \begin{pmatrix} 1 & 0{,}6 & 0{,}3 \\ 0{,}6 & 1 & 0{,}5 \\ 0{,}3 & 0{,}5 & 1 \end{pmatrix}.$$

За декомпозицією Холеского отримуємо верхню трекутню матрицю **U**

$$\mathbf{U} = \begin{pmatrix} 1 & 0{,}6 & 0{,}3 \\ 0 & 0{,}8 & 0{,}4 \\ 0 & 0 & 0{,}866 \end{pmatrix}.$$

За формулою (3) визначаємо значення трьох гаусівських випадкових величин z_1, z_2 і z_3 та формуємо вектор z

$$\mathbf{z} = (z_1, z_2, z_3)\,. \tag{4}$$

Після чого за (2) отримуємо значення $\mathbf{z_C}$ трьох залежних гаусівських випадкових величин

$$\mathbf{z_C} = (z_1, z_2, z_3)\begin{pmatrix} 1 & 0{,}6 & 0{,}3 \\ 0 & 0{,}8 & 0{,}4 \\ 0 & 0 & 0{,}866 \end{pmatrix}. \tag{5}$$

Дії (3)-(5) продовжуємо стільки разів, скільки значень залежних гаусівських випадкових величин потрібно отримати. Для цього прикладу було виконано моделювання по 2000 значень для кожної з трьох залежних гаусівських випадкових величин. При визначенні значень трьох гаусівських випадкових величин z_1, z_2 і z_3 за формулою (3) для знаходження значень u використовувався один алгоритмічний генератор

$$u_i = \frac{\omega_i - 32768}{32768};$$

$$\omega_i = (25173\omega_{i-1} + 13849) \bmod 65536.$$

Тут ω_i – i-те значення випадкової величини ω з рівномірним розподілом, $\omega \in [0, 65536)$. Початкові значення ω при моделюванні z_1, z_2 і z_3 відповідно дорівнювали 12048, 512 і 40960.

За результатами моделювання було оцінено кореляційну матрицю

$$\hat{C} = \begin{pmatrix} 1{,}002 & 0{,}592 & 0{,}295 \\ 0{,}592 & 0{,}992 & 0{,}499 \\ 0{,}295 & 0{,}499 & 1{,}009 \end{pmatrix}.$$

Максимальна відносна похибка склала 1,7% для елемента C_{13}.

Були побудовані емпіричні розподіли трьох залежних гаусівських випадкових величин z_{C_1}, z_{C_2} і z_{C_3}, для яких оцінені асиметрія та ексцес. Значення асиметрії для z_{C_1}, z_{C_2} і z_{C_3} були такими: –0,0195; 0,0708 та 0,0502, а значення ексцесу – 3,028; 3,033 та 2,974. Значення асиметрії та ексцесу для z_{C_1}, z_{C_2} і z_{C_3} практично такі ж, як для нормального розподілу. За критерієм Пірсона (χ^2) для z_{C_1}, z_{C_2} і z_{C_3} було перевірено гіпотези про нормальність їх розподілів. Значення χ^2 для z_{C_1}, z_{C_2} і z_{C_3} були такими: 98,7; 11,3 та 10,4. Критичне значення χ^2 для рівня значимості 0,05 та кількості ступенів вільності 8 дорівнює 15,5. Останнє вказує на те, що гіпотезу про нормальність розподілу для z_{C_1} за критерієм Пірсона потрібно відхилити з довірчою ймовірністю 0,95. Велике значення χ^2 для z_{C_1} пояснюється проблемою хвостів розподілу (значним розходженням з розподілом Гаусу для першого та останнього інтервалів гістограми). Виправити ситуацію на хвостах вдається за рахунок формування значень z_1 як суми двох незалежних гаусівських величин. Після такої процедури всі значення χ^2 для z_{C_1}, z_{C_2} і z_{C_3} стали менше критичного: 11,3; 14,8 і 8,2.

Отримані результати свідчать про працездатність удосконаленого метода моделювання залежних гаусівських випадкових величин на основі декомпозиції кореляційної матриці за рахунок використання нормалізуючого перетворення Джонсона із сім'ї S_B,

ВИСНОВКИ

Удосконалено метод моделювання залежних гаусівських випадкових величин на основі декомпозиції кореляційної матриці за рахунок використання нормалізуючого перетворення Джонсона із сім'ї S_B, що дозволяє зменшити витрати ресурсів комп'ютера на відповідне моделювання та максимально використати наявні можливості генераторів псевдовипадкових чисел з рівномірним розподілом за рахунок того, що для створення одного значення незалежної гаусівської випадкової величини необхідне тільки одне значення випадкової величини з рівномірним розподілом. В подальшому планується розробка нового генератора залежних гаусівських випадкових величин із використанням нормалізуючого перетворення Джонсона із сім'ї S_U.

ЛИТЕРАТУРА:

1. Shanbhag D.N. Stochastic Processes: Modeling and Simulation (Handbook of Statistics 21) /D. N. Shanbhag, C. R. Rao. – Elsevier Publishing, 2003. – 1020 p. – ISBN: 0444500138.
2. Hida T. An Innovation Approach to Random Fields: Application of White Noise Theory /T Hida, S.Si. – Publisher: World Scientific Publishing Company, 2004. – 204 p. – ISBN: 9812380957.
3. Kulkarni V.G. Introduction to Modeling and Analysis of Stochastic Systems – Second edition /V. G. Kulkarni. – Publisher: Springer, 2010. – 350 p. – ISBN: 1441917713.
4. Johnson M.E. Multivariate Statistical Simulation /M.E. Johnson. – John Wiley & Sons Inc, New York, 1987. – 230 p. – ISBN: 0471822906.
5. Probabilistic Methods in Computers: Textbook for universities on special. Computers /A.V. Krajnikov and others, ed. A. N. Lebedev and E. A. Cherniavskii. – Moscow: Visha shkol, 1986. – 312 p. (in Russian)
6. Lebedev A.N. Probabilistic methods in engineering design: Directory /A.N. Lebedev, M.S. Kupriyanov, D.D. Nedosekin, E.A. Chernyavskii. – St. Petersburg: Energoatomizdat, St. Petersburg Department, 2000. – 333 p. (in Russian)
7. Nelson R.B. An Introduction to Copulas /R.B. Nelson. – Springer, 2006. – 270 p. – ISBN: 0387286594.
8. Niaki S.T.A. Generating Correlation Matrices for Normal Random Vectors in NORTA Algorithm Using Artificial Neural Networks /S.T.A. Niaki, B. Abbasi //Journal of Uncertain Systems. – 2008. – Vol.2, No.3. – P.192-201.
9. Thijs van den Berg. Generating Correlated Random Numbers [E-resource] /Thijs van den Berg. – 2012. – Access mode: http://www.sitmo.com/article/generating-correlated-random-numbers/
10. Forsythe G.E. Computer Methods for Mathematical Computations /G.E. Forsythe, M.A. Malcolm, C.R. Moler. – Moscow: Mir, 1980. – 279 p. (in Russian)
11. Prykhodko S.B. Modelling of Gaussian Random Variables Based on Johnson Transformation from S_B Family /S.B. Prykhodko //Informatics and Mathematical Methods in Simulation. – 2012. – v.2, №1. – P.64-69. (in Ukrainian)

Рецензент: *д.т.н., проф. Коваленко І.І.,*
Національний університет кораблебудування ім. адм. Макарова, Миколаїв.