УДК 621.37

Г.В. Соколовська, м. Київ

СТАТИСТИЧНИЙ АНАЛІЗ ГЕНЕРАТОРІВ ПСЕВДОВИПАДКОЇ ПОСЛІДОВНОСТІ У ПРОГРАМНИХ СЕРЕДОВИЩАХ MATLAB ТА MATHCAD

In this paper the statistical analysis of the generators of pseudorandom sequences of programming environments Matlab and Mathcad. Analysis is based on criteria such as Pearson, Kolmogorov-Smirnov , Mises and correlation test.

Keywords: pseudo-random sequence generator, the Pearson criterion, the Kolmogorov-Smirnov criterion, Mises criterion, correlation test, Matlab, Mathcad.

Вступ. При дослідженнях природних явищ, режимів роботи технічних систем, технологічних процесів у різних галузях народного господарства, економічних процесів широко використовуються ймовірнісні моделі. Це, в першу чергу, випадкові процеси з різними законами розподілу, які відображають реальні досліджувані процеси. Комп'ютерному моделюванню реалізацій таких процесів присвячена значна кількість публікацій, в тому числі [1-4].

Відомо, що при комп'ютерному моделюванні реалізацій випадкових сигналів використовуються псевдовипадкові послідовності. Дана робота буде присвячена одній з конкретних задач, а саме: провести статистичний аналіз генераторів псевдовипадкової послідовності у програмних середовищах Matlab і Mathcad.

Необхідність постановки такої задачі пов'язана з тим, що, програмні середовища використовують відповідні програмні генератори для формування псевдовипадкової послідовності, але їх статистичні характеристики не регламентуються і не вказані в літературі.

Постановка завдання. Провести статистичний аналіз генераторів псевдовипадкової послідовності з рівномірним законом розподілу за допомогою декількох критеріїв: генератора псевдовипадкових чисел у програмному середовищі Matlab, який задається функцією rand(n,x), генератора програмного середовища Mathcad, який задається функцією rnd(x). Перевірку провоести для різних об'ємів вибірки: для n=100, n=1000, n=10000.

Основні критерії для перевірки псевдовипадкової послідовності. Відомо, що інтегрально статистичні критерії перевірки наведено в [1], а саме:

1) основні критерії перевірки псевдовипадкової послідовності:

а) критерій «хі-квадрат» (χ^2 -критерій);

б) критерій Колмогорова-Смірнова;

2) емпіричні критерії: критерій рівномірності (критерій частот); критерій

серій; критерій інтервалів; покер-критерій; критерій збирання купонів; критерій перестановок; критерій монотонності; критерій «максимум-t»; критерій конфліктів; критерій проміжків між днями народження; критерій кореляції; критерій послідовностей;

3) теоретичний критерій;

4) спектральний критерій.

Використання всіх вказаних 16 критеріїв, подальше узгодження результатів їх застосування на практиці майже не зустрічається. Це пов'язано з тим фактом, що виникають труднощі прийняття однозначного рішення. Тому в більшості випадків використовуються наступні критерії перевірки псевдовипадкової послідовності: Пірсона, Колмогорова-Смірнова, Мізеса та тест кореляції, які підтверджені практикою їх застосування.

При комп'ютерному моделюванні формується псевдовипадкова послідовність

$$X=\{x_1,\ldots,x_n\} \qquad (1)$$

як послідовність реалізацій випадкової величини ξ з заданим законом розподілу шляхом n-кратного повторного використання алгоритму моделювання.

Відомо, що для перевірки характеристик псевдовипадкової величини використовується метод перевірки статистичних гіпотез. Формулюється основна гіпотеза H_0 відносно досліджуваної характеристики, і відповідно альтернативна гіпотеза H_1. Шляхом статистичної обробки (1) і використання відповідного статистичного критерію згоди проводиться перевірка гіпотез H_0 і H_1. В даній роботі досліджуються функції розподілу відповідної випадкової величини. Необхідно за допомогою статистики (1) перевірити гіпотезу H_0 про те, що функція розподілу випадкової величини $F_e(x)=F_0(x)$, де $F_0(x)$ – теоретична функція розподілу, а $F_e(x)$ - емпірична функція розподілу. Конкуруюча гіпотеза H_1: $F_\xi(x)\neq F_0(x)$.

Більш конкретно зупинимось на змісті роботи, використовуючи вказані вище статистичні критерії згоди.

χ^2*-критерій.* Цей критерій також називають критерієм Пірсона. В (1) знаходимо $x\_=min\{x_i\}$, $x_+=max\{x_i\}$ та здійснюємо розбиття варіаційного ряду на $k>1$ комірок:

$$(x\_, x\_+h),[x\_+h, x\_+2h),\ldots,[x_+-2h),(x_+-h, x_+),$$

де $h=(x_+-x\_)/k$. Рекомендується k вибирати таким чином, щоб $\min\limits_{i=1,k} np_i \geq 5$

Розраховуємо теоретичну ймовірність попадання ξ в i-тий інтервал, якщо вірна H_0:

$$p_i = \begin{cases} F_0\left(x_- + h\right), \text{ якщо } i = 1, \\ F_0\left(x_- + ih\right) - F_0\left(x_- + (i-1)h\right), \text{ якщо } 1 < i < k, \\ 1 - F_0\left(x_+ - h\right), \text{ якщо } i = k, \end{cases} \tag{1}$$

число n_i вибіркових значень з (1), які потрапили в i-тий інтервал. При чому $\sum_{i=1}^{k} p_i = 1, \sum_{i=1}^{k} n_i = n.$

Обчислимо χ^2 -статистику:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} . \tag{2}$$

Вирішальне правило: приймається гіпотеза

$$\begin{cases} H_0, \text{ якщо } \chi^2 < \Delta, \\ H_1, \text{якщо } \chi^2 \geq \Delta, \end{cases} \tag{3}$$

де поріг Δ вибирається таким, щоб імовірність помилки першого роду рівнялася заданому рівню значимості

$$P\left\{\chi^2 \geq \Delta / H_0\right\} = \varepsilon_0 . \tag{4}$$

Критерій Колмогорова-Смірнова. Позначимо впорядкований в порядку зростання ряд вибіркових значень з X [3]: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}, n \geq 3$.

Тоді емпірична функція розподілу

$$F_e(x) = \begin{cases} 0, \text{якщо } x \leq x_{(1)}, \\ \dfrac{i}{n}, \text{ якщо } x_{(i)} < x \leq x_{(i+1)}, i = \overline{1, n-1}, \\ 1, \text{якщо } x > x_{(n)}. \end{cases} \tag{5}$$

Визначимо відстань Колмогорова між емпіричною та теоретичною функціями розподілу:

$$D = \max_x \left| F_e(x) - F_0(x) \right| . \tag{6}$$

Вирішальне правило: приймається гіпотеза

$$\begin{cases} H_0, \text{якщо } \sqrt{n} D < \Delta, \\ H_1, \text{ якщо } \sqrt{n} D \geq \Delta. \end{cases} \tag{7}$$

де поріг Δ визначається згідно (4).

Критерій ω^2. Даний критерій також називають критерій Мізеса [3], він використовує емпіричну функцію розподілу (5) та метрику типу середньоквадратичного відхилення

$$\omega^2 = \int_{-\infty}^{\infty} \left(F_0(x) - F_e(x)\right)^2 dF(x). \tag{8}$$

Статистика критерію

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F_e(x_i) - \frac{i-0{,}5}{n}\right]^2. \quad (9)$$

Вирішальне правило: приймається гіпотеза

$$\begin{cases} H_0, \text{ якщо } n\omega^2 < \Delta, \\ H_1, \text{ якщо } n\omega^2 \geq \Delta. \end{cases} \quad (10)$$

де поріг Δ визначається згідно (4).

Тест перевірки кореляції між членами послідовності. У багатьох випадках необхідно, щоб кореляція між членами послідовності псевдовипадкових чисел була незначна [4]. Для цього обчислюється коефіцієнт кореляції

$$C = \frac{n\left[\sum_{i=1}^{n-1} x_i x_{i+1} + x_n x_1\right] - \left[\sum_{i=1}^{n} x_i\right]^2}{n\sum_{i=1}^{n} x_i - \left[\sum_{i=1}^{n} x_i\right]^2}. \quad (11)$$

Вирішальне правило: приймається гіпотеза

$$\begin{cases} H_0, \text{ якщо } \mu_n - 2\sigma_n \leq C \leq \mu_n - 2\sigma_n, \\ H_1, \text{ в інших випадках.} \end{cases} \quad (12)$$

де

$$\mu_n = -\frac{1}{n-1}, \quad (13)$$

$$\sigma_n = \frac{\sqrt{n(n-3)/(n+1)}}{n-1}.$$

Основні результати. Для перевірки якості генератору псевдовипадкової послідовності програмного середовища Matlab та програмного середовища Mathcad був використаний аналіз генератора за математичним сподіванням, дисперсією та середньоквадратичним відхиленням (СКВ), а також два критерії, описані вище.

Аналіз генераторів псевдовипадкової послідовності покажемо на прикладі таблиць 1-3.

Таблиця 1

Результати аналізу генераторів псевдовипадкової послідовності для об'єму вибірки n=100

Похибки генераторів	Matlab (2011)	Mathcad (2011)
За математичним сподіванням	1.16 %	1.77 %
За дисперсією	4.57 %	0.36 %
За СКВ	2.31 %	0.18 %

Таблиця 2

Результати аналізу генераторів псевдовипадкової послідовності для об'єму вибірки n=1000

Похибки генераторів	Matlab (2011)	Mathcad (2011)
За математичним сподіванням	0,11%	0,78%
За дисперсією	2,88%	0,076%
За СКВ	1,43%	0,043%

Таблиця 3

Результати аналізу генераторів псевдовипадкової послідовності для об'єму вибірки n=10000

Похибки генераторів	Matlab (2011)	Mathcad (2011)
За математичним сподіванням	0.58 %	1.12 %
За дисперсією	3.28 %	2.04 %
За СКВ	1.65 %	1.02 %

Як видно з результатів таблиць 1-3, похибка за математичним сподіванням менша у програмному середовищі Matlab, але за двома іншими – у середовищі Mathcad. Також треба відмітити, що похибки суттєво відрізняються для різних об'ємів вибірки, а також те, що найменшими вони є для вибірки n=1000, що дає змогу зробити висновок про доцільність використання саме такого об'єму.

Оцінка якості генераторів псевдовипадкової послідовності на основі критеріїв згоди. Для цього ми брали, також, як і для аналізу генераторів, 3 різних об'єми вибірки: n=100; n=1000; n=10000. Вони перевірялися за критеріями Пірсона та Колмогорова-Смірнова, Мізеса та тестом кореляції у середовищах Matlab та Mathcad. Гіпотези, які приймаються при цьому, мали однакові значення для всіх об'ємів вибірки та проілюстровані у таблиці 4.

Таблиця 4

Гіпотези, які приймаються при використанні даних критеріїв

Критерії згоди	Прийнята гіпотеза	
	Matlab (2011)	Mathcad (2011)
Критерій Пірсона	H_0	H_1
Критерій Колмогорова-Смірнова	H_0	H_0
Критерій Мізеса	H_0	H_0
Тест кореляції	H_1	H_0

Як видно з таблиці 4, незалежно від об'єму вибірки при використанні критеріїв Колмогорова-Смірнова та Мізеса обидва генератори дають однаковий результат, але при використанні критеріїв Пірсона та тесту кореляції гіпотези приймають різне значення. При чому згідно з критерієм Пірсона, генератор псевдовипадкової послідовності у Matlab кращій, а при

використанні тесту кореляції – навпаки.

Висновки. В роботі був наведений статистичний аналіз генераторів псевдовипадкової послідовності, яка розподілена за рівномірним законом на основі використання програмних середовищ Matlab та Mathcad. Дана перевірка була представлена на основі порівняльного аналізу генераторів у двох програмних середовищах на основі оцінки похибок за математичним сподіванням, дисперсією та середньоквадратичним відхиленням, а також на основі критеріїв Колмогорова-Смірнова, Пірсона, Мізеса та тесту кореляції. Для оцінки генераторів за даними критеріями були використані об'єми вибірок n=100; n=1000; n=10000. Для кожного об'єму наведені числові результати похибок та проведений їх аналіз.

1. *Кнут Д.Э.* Искусство программирования. Том 2. Получисленные алгоритмы. – М.Мир, 2000. – 768 с.

2. *Харин Ю.С., Степанова М.Д.* Практикум на ЭВМ по математической статистике. Для мат. спец. ун-тов. – Мн.: изд-во «Университетское», 1987. – 304с.: ил.

3. *Черкесов Г.Н.* Надежность апаратно-программных комплексов. Учебное пособие. – СПб.: Питер, 2005. – 479 с.: ил.

4. *Гришин В.К.* Математическая обработка и интерпретация физического эксперимента / В.К. Гришин, Ф.А. Живописцев, В.А. Иванов. – М.: Изд-во Моск. ун-та, 1988. – 318 с.

Поступила 4.03.2013р.

УДК 681.3

С.М. Головань, А.М. Давиденко, Л.М. Щербак, м. Київ

ПРО ТЕРМІНОЛОГІЮ В ОБЛАСТІ БЕЗПЕКИ ІНФОРМАЦІЇ

In this paper, based on the results of comparative analysis of terminological framework in the area of information security today proposed a number of definitions.

Keywords: information security, information sources, threats, information security, security policy.

Вступ. Дана робота є дискусійною і присвячена питанням термінологічної бази в області безпеки інформації. Відмітимо, що по суті епіграфом до даної роботи може бути афоризм відомого французького філософа і математика Рене Декарта (1590-1650): «Визначте зміст слів і ви звільните людство від половини його непорозумінь».

Термінологічна база змінюється і вдосконалюється по мірі розвитку земної цивілізації. Це в повній мірі відноситься і до області безпеки інформації. Цій темі присвячена значна кількість публікацій, в тому числі [1, …,7].