УДК 004.4`2

Коцюба А. Ю., канд. фіз.-мат. наук, доцент (Тел.: +380 99 197 90 89. E-mail : akaerKAJ@rambler.ru)
(Луцький національний технічний університет)

ПРО МЕТОДИКУ ОЦІНЮВАННЯ ПОХИБКИ ГЕНЕРАТОРА ПСЕВДОВИПАДКОВИХ ЧИСЕЛ

Коцюба А. Ю. Про методику оцінювання похибки генератора псевдовипадкових чисел. На основі середовища MATLAB та розроблених у ньому рівномірного генератора псевдовипадкових чисел (ГПВЧ) rand() та функції randperm() побудовано експериментальну методику оцінювання похибки, яка базується на критерії згоди χ^2 -Пірсона. Для апробації даної методики в заданому середовищі створено та запрограмовано алгоритм, що дозволяє імітувати ГПВЧ з різними похибками, які користувач може контролювати за допомогою значення натурального параметра. Все це дало змогу проробити ряд експериментів. З одержаних результатів можна зробити висновки про ефективність методики оцінювання різних ГПВЧ (навіть власного) у різних середовищах програмування. При цьому не обов'язково прив'язуватися до вищезгаданого критерію згоди (можна використовувати і інші методи математичної статистики прийняття чи відхилення гіпотези про розподіл, який не обов'язково повинен бути рівномірним).

***Ключові слова*:** генератор псевдовипадкових чисел, MATLAB, критерій згоди χ^2 -Пірсона, математична статистика, національна лотерея «Супер Лото»

Коцюба А. Ю. О методике оценки погрешности генератора псевдослучайных чисел. На основе среды MATLAB и разработанных в нем равномерного генератора псевдослучайных чисел (ГПСЧ) rand() и функции randperm() построено экспериментальную методику оценки погрешности, которая основанная на критерии согласия χ^2 -Пирсона. Для апробации данной методики в заданной среде создан и запрограммирован алгоритм, позволяющий имитировать ГПВЧ с различными погрешностями, которые пользователь может контролировать с помощью значения натурального параметра. Все это позволило проделать ряд экспериментов. Из полученных результатов можно сделать выводы об эффективности методики оценки различных ГПВЧ (даже собственного) в различных средах программирования. При этом не обязательно привязываться к вышеупомянутому критерию согласия (можно использовать и другие методы математической статистики принятие или отклонение гипотезы о распределении, который не обязательно должен быть равномерным).

Ключевые слова: генератор псевдослучайных чисел, MATLAB, критерий согласия χ^2 -Пирсона, математическая статистика, национальная лотерея «Супер Лото"».

Kotsyuba A. Yu. On the method of estimation errors pseudo-random number generator. Based on MATLAB environment and developed it uniform generator pseudo-random number (GPRN) rand() and function randperm() built an experimental method of estimation errors based on criteria χ^2 -Pirson consent. To test this method in a given environment created and programmed algorithm that allows to simulate GPRN various errors that the user can control the importance of using natural setting. This made it possible to do a series of experiments. From the obtained results it is possible to draw conclusions about the effectiveness of different methods of assessment GPRN (even your own) in different programming environments. It is not necessarily tied to the above criteria consent (you can use other methods of mathematical statistics accepting or rejecting the hypothesis of distribution, which need not be uniform).

***Keywords*:** generator of pseudorandom numbers, environment MATLAB, criterion consent χ^2 -Pirson, mathematical statistic, national lottery "Super Lotto"

Вступ і постановка задачі. Зазвичай в моделюванні, імітаційному чи математичному, результати отримуємо за допомогою так званого генератора псевдовипадкових чисел (ГВПЧ) [1, 2]. Широке використання ГВПЧ отримали в системах кодування [3] і криптографічного захисту інформації [4]. Тут термін "псевдо" можна сприймати як похибку. Для реалізації точніших моделей необхідно позбутися цієї похибки. Але це зробити з математичної точки зору практично неможливо (можна зробити це "псевдо" значно меншим, але позбутися його взагалі неможливо). Для досягнення більшої точності програмний код, що описує генератор, необхідно робити більш громіздким (а для деяких моделей чи симуляцій це призведе до їх повільнішої роботи). В таких випадках необхідно дійти до компромісу між допустимою похибкою та швидкістю роботи програмного продукту.

Метою даної роботи є не створення власного генератора випадкових чисел (ГПВЧ), а розробка та аналіз методики, яка б дозволяла оцінювати таку похибку. Для її досягнення буде розроблено алгоритм побудови вибірки не зовсім випадкових чисел (на випадковість будуть накладені деякі умови), для якої можна буде приймати гіпотезу про рівномірність розподілу не рідше, ніж для вибірки, побудованої за допомогою відповідного ГПВЧ, або взагалі видати набір чисел, який завжди при заданій ймовірності буде задовольняти цю гіпотезу. Спочатку створимо сукупність вибірок, за допомогою запрограмованого генератора двома способами.

Нехай потрібно працювати з генератором, що генерує вибірку, яка задовольняє рівномірний розподіл. Зазвичай в різних середовищах програмування чи математичного моделювання такі ГПВЧ вже створені розробниками. Для прикладу візьмемо середовище MATLAB та розроблений у ньому генератор *rand()* [5, 6].

В статистиці існують різні методи, які дають можливість приймати чи відхиляти гіпотези (наприклад, гіпотезу про рівномірність розподілу). Одним з таких методів є критерії згоди. Використаємо один з найпростіших для розуміння критерій згоди χ^2-Пірсона. Він полягає у тому, що для вибірки спочатку обчислюється величина

$$\rho\left(\vec{X}\right)=\sum_{j=1}^{k}\frac{\left(\nu_j-np_j\right)^2}{np_j},$$

де $A_1, A_2, ..., A_k$ – інтервали групування в області значень випадкової величини (ВВ), $\vec{X}=\left(X_1, X_2, ..., X_n\right)$ – вибірка ВВ; ν_j – кількість елементів вибірки, що попали в інтервал групування A_j, $j=\overline{1,k}$; $p_1=p_2=...=p_k=1/k$, а k – кількість інтервалів групування.

А потім обчислюється на основі величини ймовірність прийняття або відкидання вищевказаної гіпотези.

Побудова та аналіз методики. Для прикладу розглянемо імітаційне моделювання симуляції розіграшів національної лотереї "Супер Лото" 6 з 52, тобто приймемо, що $k=52$ (кожен інтервал групування, це номер кульки). Цей приклад є цікавим ще й тому, що вибірку можна будувати двояко:

– по-перше, можна робити вибірку, не зважаючи на симуляцію, тобто таку, що серед 6-ти перших елементів цієї вибірки можуть бути однакові;

– по-друге, робити вибірку, у якій задаємо по 6 різних номерів кульок, тобто як у симуляціях розіграшів.

При цьому цікаво було б проаналізувати різницю між такими випробуваннями.

Побудуємо для обох випадків *100* вибірок об'ємом $1200j$, де $j=\overline{1,100}$ (вибірки можна робити різних об'ємів, наприклад, замість множника j можна задавати натуральну ВВ, або взагалі задавати скрізь фіксований об'єм). Наш підхід з об'ємом дає в майбутньому можливість дослідити зв'язок між об'ємами цих вибірок та кінцевим результатом (точніше дослідити той факт, що такого зв'язку немає). Для кожної з таких вибірок розглянемо *3* випадки, у яких ймовірності прийняття гіпотези відповідно становлять 0.5, 0.4 та 0.3. Для кожного з випадків обчислимо сумарну кількість вибірок, для яких приймається гіпотеза. Отримаємо три випадкових значення. Повторимо ці міркування, наприклад, 500 разів. В результаті одержимо *3* вибірки об'ємом 500. За значеннями цих вибірок створимо інтервали групування аналогічно як для критерію згоди χ^2-Пірсона і підрахуємо для кожного з них відповідні кількості. Для більшої наочності зробимо по 4 експерименти для випадків нереальної симуляції розіграшів (Рис. 1) та реальної (Рис. 2).

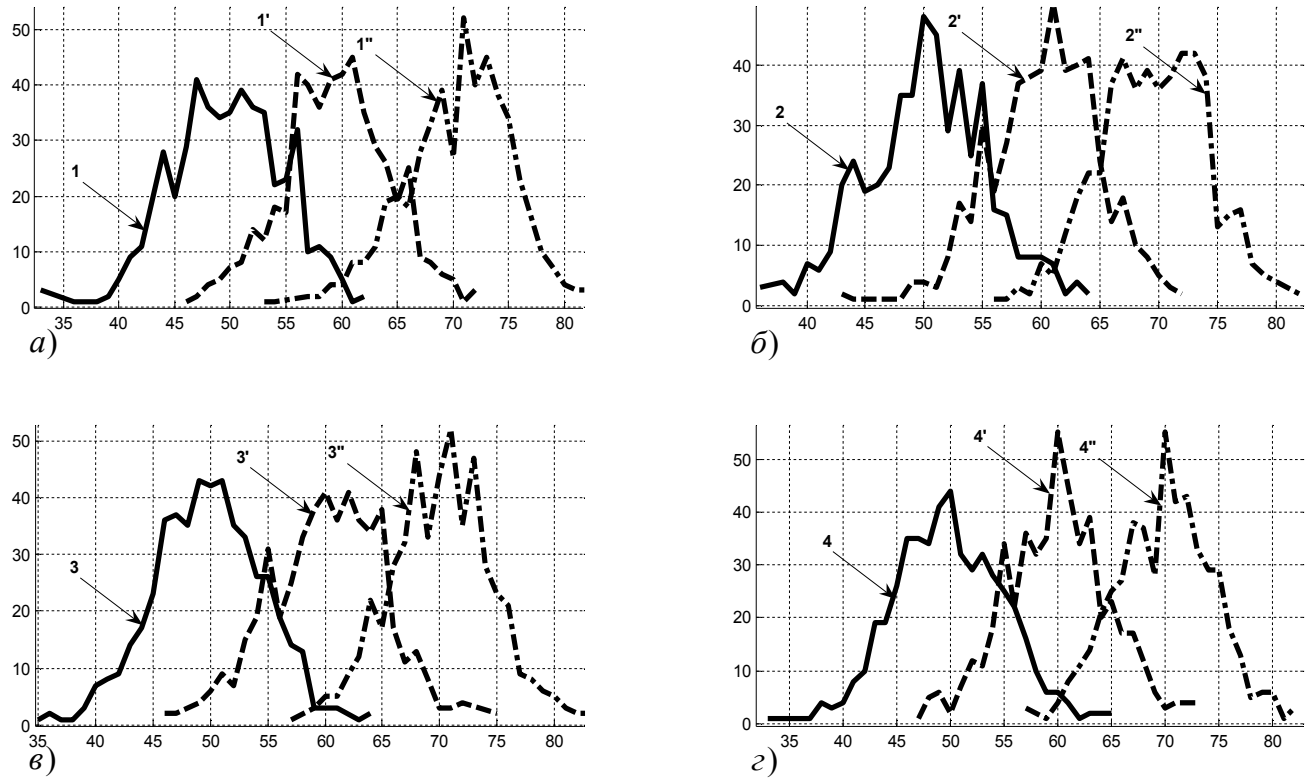

Рис. 1. Результати експериментів для нереальних симуляції розіграшів національної лотереї "Супер Лото"

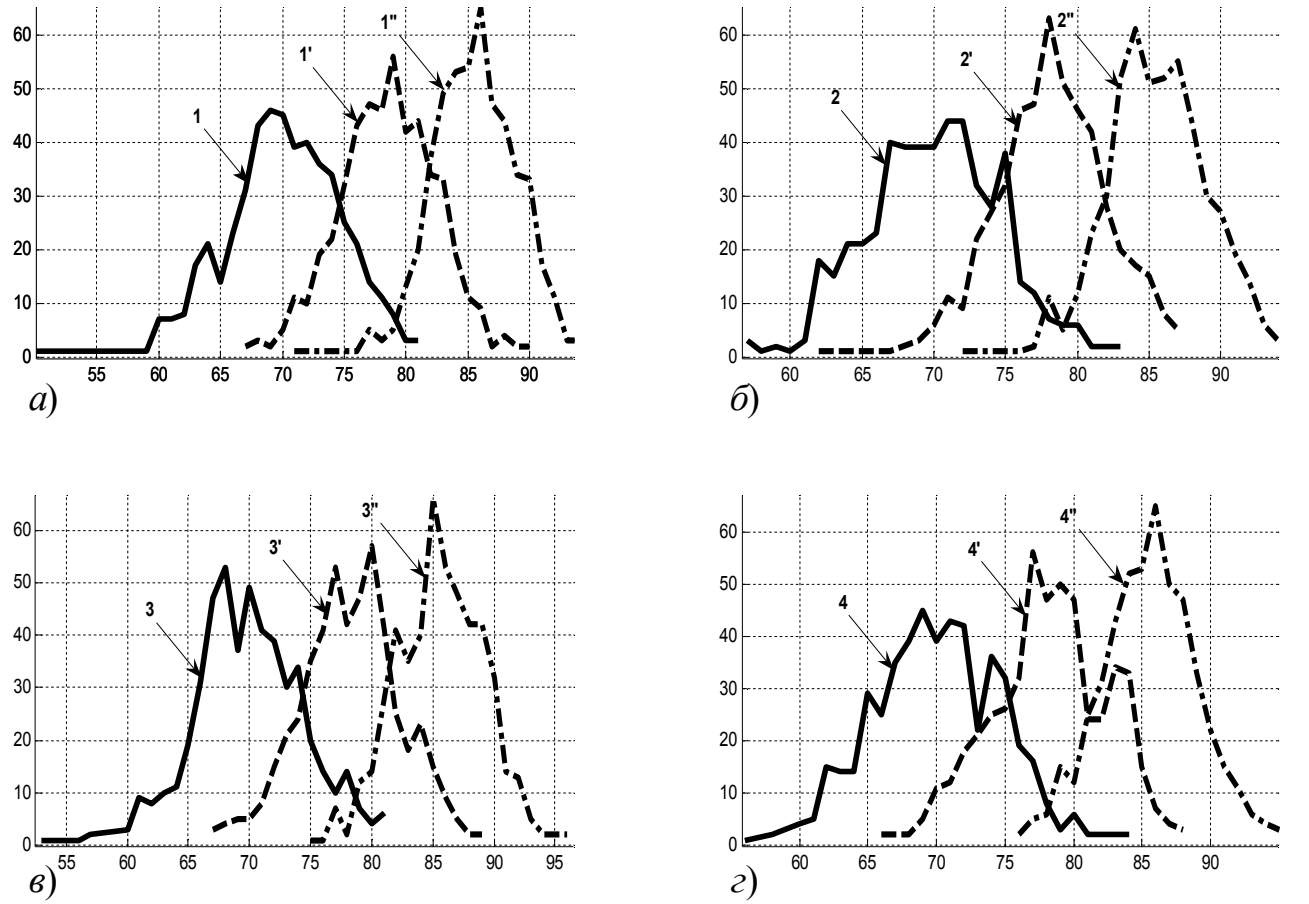

Рис. 2. Результати експериментів для реальних симуляції розіграшів національної лотереї "Супер Лото"

На рис. 1 та 2 подано результати *4*-х експериментів – а)...г) для того, щоб переконатися, що одержані відповідні результати-вибірки з великою ймовірністю задовольняють близькі за своїми характеристиками нормальні розподіли. Іншими словами, якщо побудувати методом найменших квадратів регресії нормального розподілу і за невідомі приймати такі характеристики як математичне сподівання та дисперсію, то, очевидно, не залежно від номера експерименту одержимо близькі значення цих характеристик для кожної з 4-х відповідних вибірок. Тут кривим (1...4) відповідає випадок, коли ймовірність прийняття гіпотези становить 0.5; кривим (1′...4′) – відповідає ймовірність 0.4; а кривим (1″...4″) – 0.3.

Очевидно, що відповідні математичні сподівання при зменшенні ймовірностей прийняття гіпотези зростають і в кінцевому випадку нормальний розподіл вироджується (про це виродження детальніше розповімо нижче). Крім цього, очевидно, що цей так званий процес виродження на Рис. 2 відбувається значно швидше, ніж на Рис. 1. На основі таких результатів зробимо висновок, що другим способом досягається зменшення похибки ГПВЧ.

Розроблена методика показує як аналіз вибірки, що повинна задовольняти рівномірний розподіл ВВ можна звести до аналізу вибірки, що, очевидно, задовольняє нормальний розподіл новоутвореної ВВ. Причому тут подібним чином можна використовувати як вищевказаний, так і інші критерії згоди, наприклад, Колмогорова-Смірнова тощо. А для нормального розподілу визначаючими є такі вищеописані характеристики як математичне сподівання та дисперсія або середнє квадратичне відхилення. Крім того цей підхід дає змогу показати як за допомогою генератора *rand()* можна отримати ГПВЧ, що задовольняє нормальний розподіл.

Таким чином, у даній роботі розроблена методика, що базується на аналізі характеристик регресій, отриманих з новоутвореної вибірки, дає можливість зробити висновки про ГПВЧ, що генерує псевдорівномірний розподіл. При цьому не зайвим буде збільшення об'єму вихідної вибірки (> 500). Це дасть можливість зробити точнішу оцінку похибки. Але вже навіть з отриманих результатів можна зробити висновок, що використання реальної моделі симуляції зменшує похибку генератора, тобто можна придумати алгоритм створення сукупності вибірок, для яких із вказаною ймовірністю буде більша кількість прийнятих гіпотез про рівномірність, тобто процес виродження буде ще швидшим.

Алгоритм імітації ГПВЧ з різними похибками. Ідея цього алгоритму полягає в наступному:

- крім величини k, яка для даних експериментів є відомою, введемо деяку натуральну величину m, нехай $m = 6$ (це означатиме що з будь-яких 6-ти підряд згенерованих чисел, не повинно бути однакових);
- очевидно, що проведені експерименти таким чином дадуть більш точніші результати, ніж на Рис. 2 (див. Рис. 3);
- очевидно, також, що при $m = k$, гіпотези про рівномірність завжди з великою ймовірністю будуть прийматися;
- виникає логічне питання, а що відбувається при $6 < m < k$;
- проведемо ряд експериментів, які дадуть можливість відповісти на це питання.

Застосуємо вищеописаний алгоритм для випадків $m = 6$ та $m = 11$, проведемо по 4 експерименти, результати яких зобразимо відповідно на Рис. 3 та Рис. 4.

Очевидно, що процес виродження у новоутворених за допомогою вищеописаного алгоритму вибірках відбувається значно швидше. На Рис. 3 вже при ймовірності прийняття гіпотези 0.3 зустрічаються експерименти, для яких зі 100 згенерованих рівномірним генератором вибірок приймаються всі гіпотези. На Рис. 1 та Рис. 2 для досягнення такого результату ймовірність прийняття гіпотези необхідно було б суттєво зменшити. Щодо останніх експериментів (при $m = 11$), результати яких подано на Рис. 4, то тут можна спостерігати майже кінцевий результат процесу виродження.

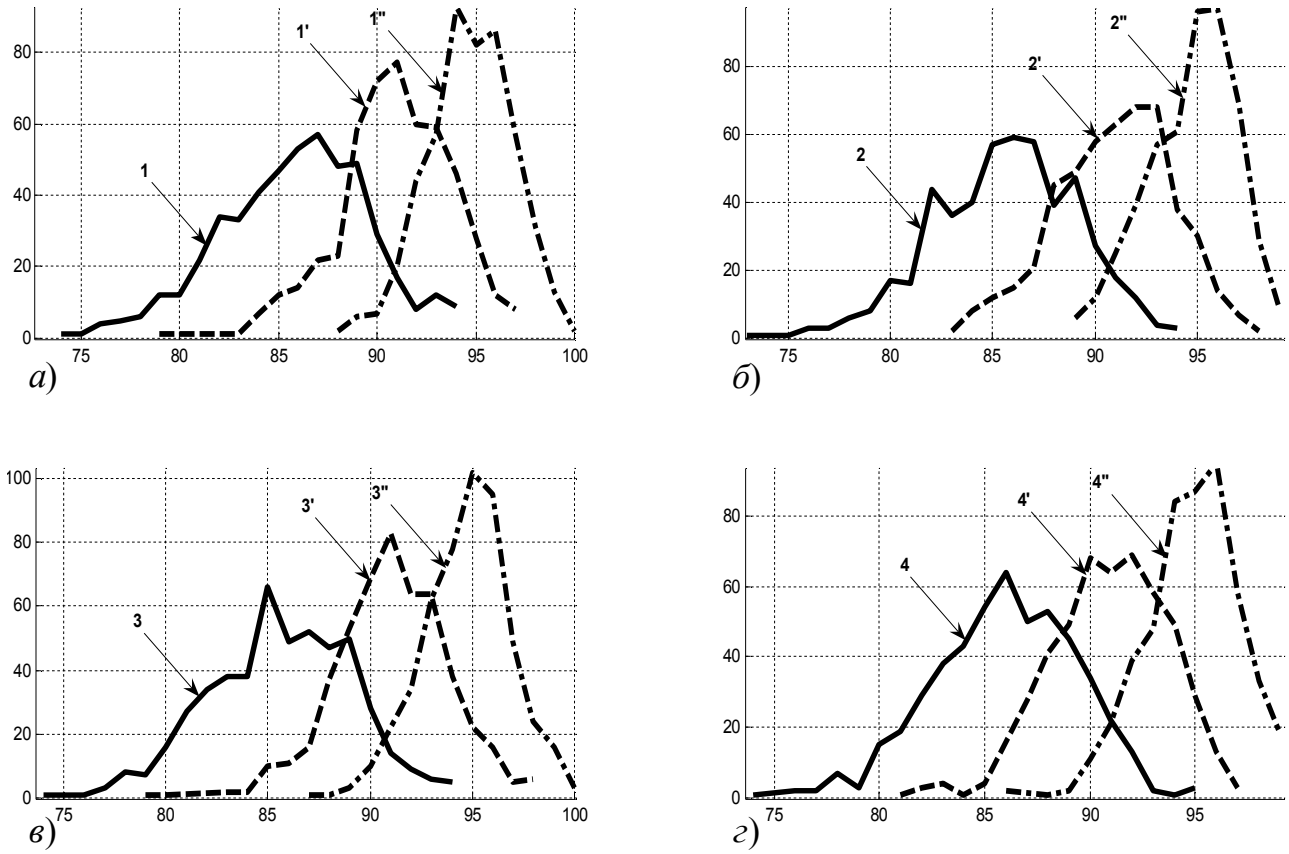

Рис. 3. Результати експериментів при $m = 6$

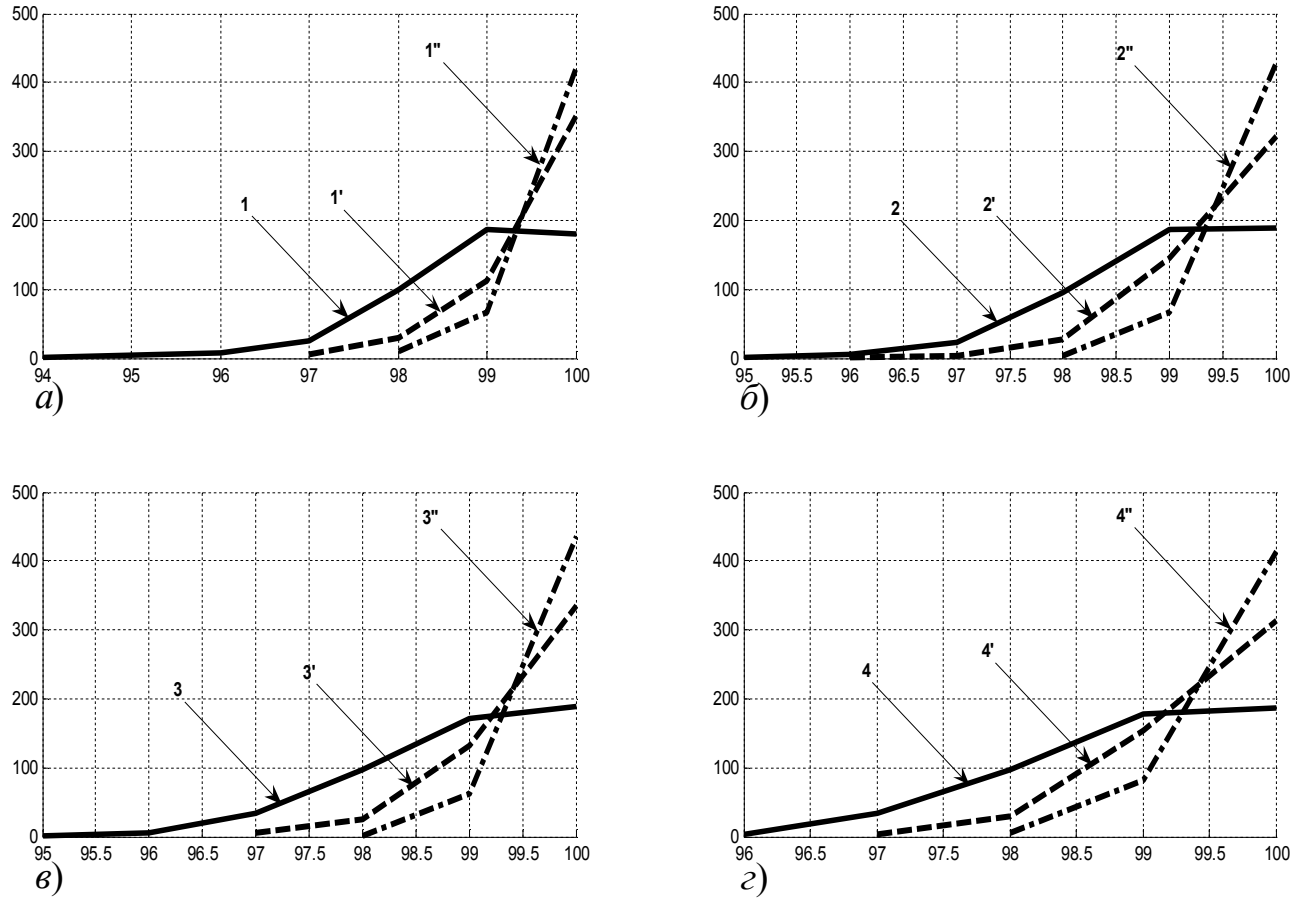

Рис. 4. Результати експериментів при $m = 11$

Очевидно, що кінцевим результатом виродження є вибірка заданого об'єму, у якій всі її значення будуть рівними кількості вибірок згенерованих рівномірним ГПВЧ, для проведених у даній роботі експериментів виродженою вибіркою-результатом буде $\underbrace{100,100,...,100}_{500}$ і графічно результат представляється точкою з координатами $(100;500)$. Таким чином методика оцінювання похибки генератора псевдовипадкових чисел (не обов'язково на рівномірність і не лише за критерієм згоди χ^2-Пірсона) полягає у визначенні такої ймовірності прийняття гіпотези про те, що утворені вибірки задовольняють вказаному розподілу, для якої вихідна або новоутворена вибірка-результат буде завжди виродженою.

Висновки. З експериментів зображених відповідно на Рис. 1...4 можна зробити наступні висновки:

– для рівномірного ГПВЧ *rand()* в середовищі MATLAB така ймовірність є значно меншою за 0.3 (шукана ймовірність є $<< 0.3$ для результатів поданих на Рис. 1);

– для результатів наведених на рис. 2 шукана ймовірність хоча і є меншою за 0.3, але неважко переконатися, що становить вона ≈ 0.1 (можливо є дещо меншою);

– візуальний аналіз результатів наведених на Рис. 3 показує, що шукана ймовірність становить ≈ 0.2 (можливо вона є дещо меншою);

– для результатів наведених на Рис. 4, очевидно, що шукана ймовірність є несуттєво меншою за 0.3;

– крім того як показують експерименти, для того, щоб шукана ймовірність була не меншою за 0.3 достатньо, щоб у розробленому в даній роботі алгоритмі побудові вибірки покласти, що $m > 15$ (для будь якої фіксованої ймовірності завжди можна підібрати натуральну величину $m < k$);

– аналогічно можна стверджувати, що для будь-якого ГПВЧ можна завжди порахувати ймовірність таку, що для всіх вибірок буде прийматися гіпотеза (очевидно, що для уточнення оцінки такої ймовірності необхідно збільшувати кількість експериментів).

І хоча розроблений у даній роботі алгоритм побудови вибірки не є генератором випадкових чисел, він дав можливість зробити симуляцію роботи більш точнішого ГПВЧ і зробити відповідні висновки. Це повинно допомогти оцінювати похибки генераторів псевдовипадкових чисел у будь-якому середовищі, навіть власного ГПВЧ.

Література

1. Герасимчук О. І. Генератори псевдовипадкових чисел, їх застосування, класифікація, основні методи побудови і оцінка якості / О. І. Герасимчук, В. М. Максимович // Захист інформації. – 2003. – №3. – С. 29-36.

2. Генератор псевдовипадкових чисел [Електронний ресурс] // – Режим доступу http://uk.wikipedia.org/wiki/Генератор_псевдовипадкових_чисел. (11.05.2014 р.)

3. Семенко А. И. Создание модифицированных псевдослучайных последовательностей Голда для телекоммуникационных систем с кодовым разделением каналов / А. И. Семенко, Н. И. Бокла // Телекомунікаційні та інформаційні технології. – 2014. – №3. – С. 10-14.

4. Алгоритм шифрування інформації з використанням псевдовипадкових послідовностей / О. В. Гресь, Р. Л. Політанський, П. М. Шпатар, А. Д. Верига // Наукові записки Українського науково-дослідного інституту зв'язку. – 2013. – №1(25). – С. 88-93.

5. Ануфриев И. Е. MATLAB 7 / И. Е. Ануфриев, А. Б. Смирнов, Е. Н. Смирнова – Санкт-Петербург : БХВ-Петербург, 2005. – 1104 с.

6. Кетков Ю.Л. MATLAB 7: Программирование, численные методы / Ю. Л. Кетков, А. Ю. Кетков, М. М. Шульц – Санкт-Петербург : БХВ-Петербург, 2005. – 752 с.