Лабораторна робота №1.
Побудова імітаційної моделі генератора псевдовипадкових чисел (ГПВЧ).
Перевірка якості роботи генератора

Мета роботи – ознайомитися з еталоном функціонування ГПВЧ; побудувати імітаційну модель функціонування ГПВЧ на основі лінійного конгруентного методу та здійснити перевірку якості роботи створеного ГПВЧ.

Короткі теоретичні відомості

Генератори випадкових чисел (ГВЧ)

Математичне сподівання m_x та дисперсія D_x такої послідовності, що складається з n випадкових чисел x_i, повинні бути такими (якщо це дійсно рівномірно розподілені випадкові числа в інтервалі від 0 до 1):

$$m_x = \frac{\sum_{i=1}^{n} x_i}{n} = 0.5 \qquad (1.1)$$

Якщо користувачеві необхідно, щоб випадкове число x' знаходилося в інтервалі $(a; b)$, відмінному від (0; 1), потрібно скористатися формулою $x' = a + (b - a)x$, де x — випадкове число з інтервалу (0; 1). Тепер x' — випадкове число, рівномірно розподілене в діапазоні від a до b.

За еталон ГВЧ прийнято такий генератор, який породжує послідовність випадкових чисел за рівномірним законом розподілу в інтервалі (0; 1). За одне звернення даний генератор повертає одне випадкове число. Якщо спостерігати такий ГВЧ досить тривалий час, то виявиться, що, наприклад, в кожен із десяти інтервалів (0; 0.1), (0.1; 0.2), (0.2; 0.3), ..., (0.9; 1) потрапить практично однакова кількість випадкових чисел – тобто вони будуть розподілені рівномірно по всьому інтервалу (0; 1). Якщо зобразити на графіку $m=10$ інтервалів і частоти N_i влучень в них, то вийде експериментальна крива щільності розподілу випадкових чисел (рис. 1.1).

Зауважимо, що в ідеалі крива щільності розподілу випадкових чисел виглядала б так, як показано на рис. 1.2. Тобто в ідеальному випадку в кожен інтервал потрапляє однакове число точок: $N_i = N/m$, де N — загальне число точок, m — кількість інтервалів, $i = \overline{1,m}$.

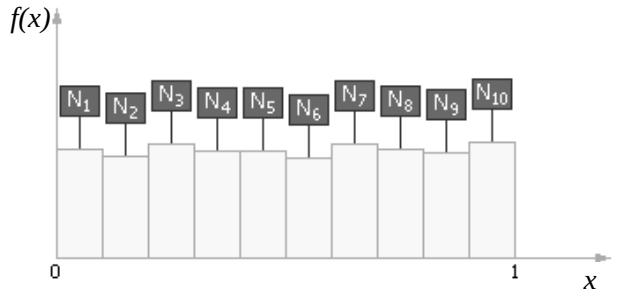

Рис. 1.1. Частотна діаграма випадання випадкових чисел, породжуваних реальним генератором

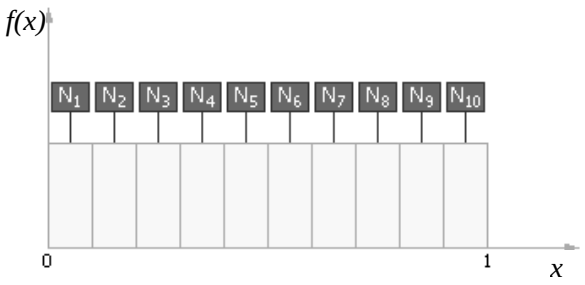

Рис. 1.2. Частотна діаграма випадання випадкових чисел, що породжуються ідеальним генератором теоретично

Відмітимо, що генерація довільного випадкового числа складається з двох етапів:

- генерація нормалізованого випадкового числа (тобто рівномірно розподіленого від 0 до 1);
- перетворення нормалізованих випадкових чисел X_i в випадкові числа x_i, які розподілені за необхідним користувачеві (довільному) законом розподілу або в необхідному інтервалі.

Генератори випадкових чисел за способом отримання чисел діляться на фізичні, табличні та алгоритмічні.

Алгоритмічні генератори псевдовипадкових чисел (ГПВЧ)

Числа, які генеруються за допомогою цих ГПВЧ, завжди є псевдовипадковими (або квазівипадковими), тобто кожне наступне згенероване число залежить від попереднього: $x_{i+1} = f(x_i)$.

Послідовності, складені з таких чисел, утворюють петлі, тобто обов'язково існує цикл, що повторюється нескінченну кількість раз. Повторювані цикли називаються періодами.

Перевагою даних ГПВЧ є швидкодія; генератори практично не вимагають ресурсів пам'яті, компактні. Недоліки: числа неможна в повній мірі назвати випадковими, оскільки між ними існує залежність, а також наявність періодів в послідовності квазівипадкових чисел.

Лінійний конгруентний метод (мультиплікативний метод)

Поширений метод для генерації псевдовипадкових чисел, що не володіє криптографічною стійкістю. Лінійний конгруентний метод полягає в обчисленні членів лінійної рекурентної послідовності по модулю деякого натурального числа *M*, що задається такою формулою:

$$X_{i+1} = (\lambda\, X_i + \mu) \bmod M, \tag{1.2}$$

де $M = q^n - 1$; q – основа системи числення, прийнятої в комп'ютері; n – кількість цифрових розрядів у машинному слові; $\lambda = 8\alpha \pm 3$, α – ціле додатне число; μ – ціле додатне непарне число; X_0 – ціле додатне непарне число.

Чергове число x_{i+1} псевдовипадкової послідовності обчислюється в результаті нормалізації цілого числа X_{i+1} за формулою:

$$x_{i+1} = \frac{X_{i+1}}{M}. \tag{1.3}$$

Отримана послідовність залежить від вибору початкового числа X_0 та при різних його значеннях виходять різні послідовності псевдовипадкових чисел.

Перевірка якості роботи генератора

Від якості роботи ГПВЧ залежить якість роботи всієї системи та точність результатів. Тому випадкова послідовність, породжувана ГПВЧ, повинна задовольняти цілому ряду критеріїв. Здійснювані перевірки бувають двох типів: перевірки на рівномірність розподілу та перевірки на статистичну незалежність.

Перевірка на рівномірність розподілу

1) *ГПВЧ повинен видавати близькі до наступних значення статистичних параметрів, характерних для рівномірного випадкового закону*:

$m_x = \dfrac{\sum_{i=1}^{n} x_i}{n} \approx 0.5$ — математичне очікування, $D_x = \dfrac{\sum_{i=1}^{n} (x_i - m_x)^2}{n} \approx 0.0833$ — дисперсія,

$\sigma_x = \sqrt{D_x} \approx 0.2887$ — середньоквадратичне відхилення.

2) *Частотний тест*

Частотний тест дозволяє з'ясувати, скільки чисел потрапило в інтервал ($m_x - \sigma_x$; $m_x + \sigma_x$), тобто (0.5 – 0.2887; 0.5 + 0.2887) або, в кінцевому випадку, (0.2113; 0.7887). Оскільки 0.7887 – 0.2113 = 0.5774, робимо висновок, що в гарному ГПВЧ в цей інтервал має потрапляти близько 57.7% з усіх випавших випадкових чисел (рис. 1.3).

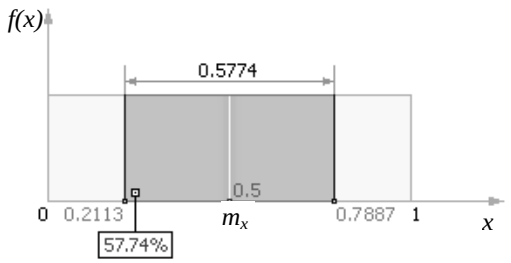

Рис. 1.3. Частотна діаграма ідеального ГПВЧ у випадку перевірки його на частотний тест

Також необхідно враховувати, що кількість чисел, які потрапили в інтервал (0; 0.5), має бути приблизно дорівнювати кількості чисел, що потрапили в інтервал (0.5; 1).

3) *Перевірка за критерієм «хі-квадрат»*

Критерій «хі-квадрат» (χ^2- критерій) – це один з найвідоміших статистичних критеріїв; він є основним методом, використовуваним в поєднанні з іншими критеріями.

Для нашого випадку перевірка за критерієм «хі-квадрат» дозволить дізнатися, наскільки створений нами реальний ГПВЧ близький до еталону ГПВЧ, тобто чи задовольняє він вимогу рівномірного розподілу.

Оскільки закон розподілу еталонного ГПВЧ рівномірний, то (теоретична) ймовірність p_j попадання чисел в j-й інтервал (всього цих інтервалів m на (0, 1)) дорівнює $p_j=1/m$ ($j=\overline{1,m}$). Таким чином, в кожен із m інтервалів потрапить рівно по $N_j^* = p_j N$ чисел, де N — кількість псевдовипадкових чисел x_i у перевіреній послідовності.

Отже, N_j — емпірична частота попадання чисел x_i в кожен j-й інтервал, тобто кількість чисел x_i, що належать j-му інтервалу. N_j^* — теоретична частота попадання чисел послідовності $(x_i;\ i=\overline{1,N})$ в m інтервалів.

Гіпотеза про рівномірність розподілу чисел в інтервалі $(0, 1)$ перевіряється за допомогою критерія Пірсона. Для цього обчислюється значення випадкової величини:

$$\chi^2{}_{експ} = \sum_{j=1}^{m} \frac{(N_j - N_j^*)^2}{N_j^*} = \frac{m}{N} \sum_{j=1}^{m} \left(N_j - \frac{N}{m} \right)^2 \qquad (1.4)$$

В табл. 1.1 наведено теоретичні значення «хі-квадрат» ($\chi^2_{теор.}$), де $v = m - 1$ — це число ступенів свободи, p — це довірча ймовірність, що задається користувачем, який вказує, наскільки ГПВЧ повинен задовольняти вимогам рівномірного розподілу, або p — це ймовірність того, що експериментальне значення $\chi^2_{експ.} \le \chi^2_{теор.}$.

Таблиця 1.1. Деякі процентні точки χ^2-розподілу

	p = 1%	**p = 5%**	**p = 25%**	**p = 50%**	**p = 75%**	**p = 95%**	**p = 99%**
$v = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$v = 2$	0.02010	0.1026	0.5754	1.386	2.773	5.991	9.210
$v = 3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$v = 4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$v = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$v = 6$	0.8721	1.635	3.455	5.348	7.841	12.59	16.81
$v = 7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$v = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$v = 9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$v = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$v = 11$	3.053	4.575	7.584	10.34	13.70	19.68	24.72
$v = 12$	3.571	5.226	8.438	11.34	14.85	21.03	26.22
$v = 15$	5.229	7.261	11.04	14.34	18.25	25.00	30.58
$v = 20$	8.260	10.85	15.45	19.34	23.83	31.41	37.57
$v = 30$	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$v = 50$	29.71	34.76	42.94	49.33	56.33	67.50	76.15
$v > 30$	$v + \text{sqrt}(2v) \cdot x_p + 2/3 \cdot x^2_p - 2/3 + O(1/\text{sqrt}(v))$						
$x_p =$	–2.33	–1.64	–0.674	0.00	0.674	1.64	2.33

Прийнятним вважають p від 10% до 90%.

Якщо $\chi^2_{експ.} > \chi^2_{теор.}$ (тобто p – велике), то генератор не задовольняє вимогу рівномірного розподілу, оскільки спостережувальні значення n_i занадто відрізняються від

теоретичних $p_j N$ і не можуть розглядатися як випадкові. Іншими словами, встановлюється такий великий довірчий інтервал, що обмеження на числа стають не дуже жорсткими, вимоги до чисел – слабкими. При цьому буде спостерігатися дуже велика абсолютна похибка.

Якщо $\chi^2_{\text{експ.}} < \chi^2_{\text{теор.}}$ (тобто p – мале), то генератор не задовольняє вимогу випадкового рівномірного розподілу, оскільки спостережувальні значення n_i занадто близькі до теоретичних $p_j N$ і не можуть розглядатися як випадкові.

А ось якщо $\chi^2_{\text{експ.}}$ лежить в деякому діапазоні між двома значеннями $\chi^2_{\text{теор.}}$, які відповідають, наприклад, $p = 25\%$ та $p = 50\%$, то можна вважати, що значення випадкових чисел, що породжуються генератором, цілком є випадковими.

При цьому додатково треба мати на увазі, що всі значення $p_j N$ повинні бути досить великими, наприклад більше 5 (з'ясовано емпіричним шляхом). Тільки тоді (при досить великий статистичної вибірці) умови проведення експерименту можна вважати задовільними.

Отже, процедура перевірки складається з таких етапів:

1. Діапазон від 0 до 1 розбивається на m рівних інтервалів.
2. Запускається ГПВЧ N разів (N повинно бути великим, наприклад, $N/m > 5$).
3. Визначається кількість випадкових чисел, що потрапили в кожен інтервал: N_j, $j = 1, \ldots, m$.
4. Обчислюється експериментальне значення $\chi^2_{\text{експ.}}$ за формулою (1.4).
5. Шляхом порівняння експериментально отриманого значення $\chi^2_{\text{експ.}}$ із теоретичним $\chi^2_{\text{теор.}}$ (з табл. 1.1) робиться висновок про придатність генератора для використання. Для цього: а) в табл. 1.1 (рядок=кількість експериментів-1); б) порівнюємо обчислене $\chi^2_{\text{експ.}}$ із $\chi^2_{\text{теор.}}$, що зустрічається в рядку. При цьому можливі три випадки.

Випадок 1: $\chi^2_{\text{експ.}} > \chi^2_{\text{теор.}}$ в рядку — гіпотеза про випадковість рівномірного розподілу ГПВЧ не виконується (розкид чисел занадто великий, щоб бути випадковим).

Випадок 2: $\chi^2_{\text{експ.}} < \chi^2_{\text{теор.}}$ в рядку — гіпотеза про випадковість рівномірного розподілу ГПВЧ не виконується (розкид чисел занадто малий, щоб бути випадковим).

Випадок 3: $\chi^2_{\text{експ.}}$ лежить в деякому діапазоні між двома значеннями $\chi^2_{\text{теор.}}$ двох сусідніх стовпців — гіпотеза випадковість рівномірного розподілу ГПВЧ виконується з імовірністю p (тобто в р випадках із 100).

Зауважимо, що чим ближче виходить p до значення 50%, тим краще.

Завдання для самостійного виконання

1. Побудувати імітаційну модель, яка відображає роботу генератора псевдовипадкових чисел. Отримати послідовність ПВЧ на інтервалі (0; 1). Використати мультиплікативний метод.
2. Виконати перевірку якості роботи генератора на рівномірність розподілу ПВЧ на інтервалі (0; 1), використовуючи критерій Пірсона та частотний тест.
3. Для отриманої послідовності ПВЧ обчислити значення статистичних параметрів: математичне очікування, дисперсія, середньоквадратичне відхилення.
4. Побудувати імітаційну модель, яка відображає роботу генератора псевдовипадкових чисел. Отримати послідовність ПВЧ на інтервалі (a; b).