Лекція 22.

Генератори випадкових чисел

СЛУЧАЙНЫЕ ЧИСЛА

Каждый, кто использует арифметические методы генерирования случайных чисел, безусловно, грешит.

— ДЖОН ФОН НЕЙМАН (JOHN VON NEUMANN) (1951)

О вероятности коль кто забудет, обманщиком вовек не будет.

— ДЖОН ГЕЙ (JOHN GAY) (1727)

Достаточно лишь нескольких лучей света, чтобы помочь людям в совершенствовании их "стохастических" способностей.

— ДЖОН ОУЭН (JOHN OWEN) (1662)

В основі методу Монте-Карло (див. Лекцію 21) лежить генерація випадкових чисел, які повинні бути рівномірно розподілені в інтервалі (0; 1).

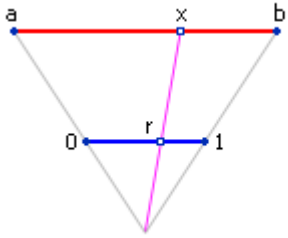
Якщо генератор видає числа, зміщені в якусь частину інтервалу (одні числа випадають частіше інших), то результат рішення задачі, розв'язуваної статистичним методом, може виявитися невірним. Тому проблема використання гарного генератора дійсно випадкових і дійсно рівномірно розподілених чисел коштує дуже гостро.

Математичне очікування m_r і дисперсія D_r такої послідовності, що складає з n випадкових чисел r_i, повинні бути наступними (якщо це дійсно рівномірно розподілені випадкові числа в інтервалі від 0 до 1):

$$m_r = \frac{\sum_{i=1}^{n} r_i}{n} = 0.5$$

$$D_r = \frac{\sum_{i=1}^{n} (r_i - m_r)^2}{n} = \frac{1}{12}$$

Якщо користувачеві буде потрібно, щоб випадкове число x перебувало в інтервалі $(a; b)$, відмінному від $(0; 1)$, потрібно скористатися формулою $x = a + (b - a) \cdot r$, де r — випадкове число з інтервалу $(0; 1)$. Законність даного перетворення демонструється на мал. 22.1.



$$\frac{r-0}{1-0}=\frac{x-a}{b-a} \Leftrightarrow r=\frac{x-a}{b-a} \Leftrightarrow x=a+(b-a)\cdot r$$

Рис. 22.1. Схема перекладу числа з інтервалу (0; 1) в інтервал (a; b)

Тепер x — випадкове число, рівномірно розподілене в діапазоні від a до b.

За **еталон генератора випадкових чисел** (ГВЧ) прийнятий такий генератор, що породжує **послідовність** випадкових чисел з *рівномірним* законом розподілу в інтервалі (0; 1). За один обіг даний генератор повертає одне випадкове число. Якщо спостерігати такий ГВЧ досить тривалий час, то виявиться, що, наприклад, у кожний з 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 інтервалі (0; 1). Якщо зобразити на графіку $k = 10$ інтервалів і частоти N_i влучень у них, то вийде експериментальна крива щільності розподілу випадкових чисел (див. мал. 22.2).

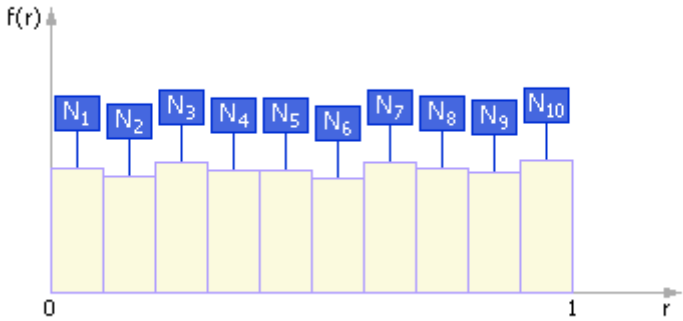


Рис. 22.2. Частотна діаграма випадання випадкових чисел, породжуваних реальним генератором

Помітимо, що в ідеалі крива щільності розподілу випадкових чисел виглядала б так, як показано на мал. 22.3. Тобто в ідеальному випадку в кожен інтервал попадає однакове число крапок: $N_i = N/k$, де N — загальне число крапок, k — кількість інтервалів, $i = 1, \ldots, k$...

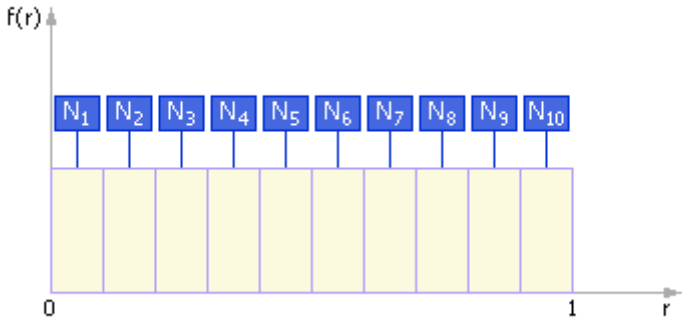


Рис. 22.3. Частотна діаграма випадання випадкових чисел, породжуваних ідеальним генератором теоретично

Варто пам'ятати, що генерація довільного випадкового числа складається із двох етапів:

- генерація нормалізованого випадкового числа (тобто рівномірно розподіленого від 0 до 1);
- перетворення нормалізованих випадкових чисел r_i у випадкові числа x_i, які розподілені по необхідному користувачі (довільному) закону розподілу або в необхідному інтервалі.

Генератори випадкових чисел по способі одержання чисел діляться на:

- фізичні;
- табличні;
- алгоритмічні.

Фізичні ГВЧ

Прикладом фізичних ГВЧ можуть служити: монета («орел» — 1, «решка» — 0); грального кістки; поділений на сектори із цифрами барабан зі стрілкою; апаратурний генератор шуму (ГШ), у якості якого використають шумливий тепловий пристрій, наприклад, транзистор (мал. 22.4–22.5).

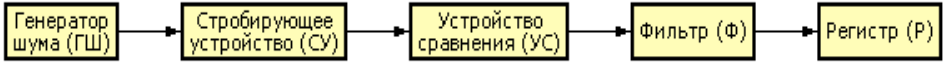


Рис. 22.4. Схема апаратного методу генерації випадкових чисел

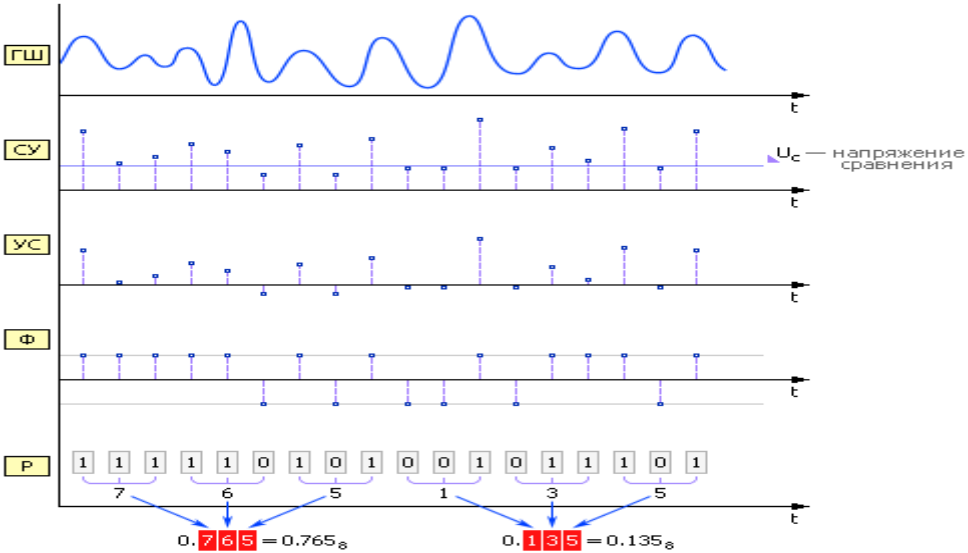


Рис. 22.5. Діаграма одержання випадкових чисел апаратним методом

Задача «Генерація випадкових чисел за допомогою монети»

Згенеруйте випадкове трирозрядне число, розподілене за рівномірним законом в інтервалі від 0 до 1, за допомогою монети. Точність - три знаки після коми.

Перший спосіб рішення задачі
Підкиньте монету 9 разів, і якщо монета впала решкою, то запишіть «0», якщо орлом, те «1». Отже, допустимо, що в результаті експерименту одержали випадкову послідовність 100110100.

Накреслите інтервал від 0 до 1. Зчитуючи числа в послідовності ліворуч праворуч, розбивайте інтервал навпіл і вибирайте щораз одну із частин чергового інтервалу (якщо випав 0, те ліву, якщо випала 1, те праву). Таким чином, можна добратися до будь-якої крапки інтервалу, як завгодно точно.

Отже, **1**: інтервал [0; 1] ділиться навпіл — [0; 0.5] й [0.5; 1], — вибирається права половина, інтервал звужується: [0.5; 1]. Наступне число, **0**: інтервал [0.5; 1] ділиться навпіл — [0.5; 0.75] й [0.75; 1], — вибирається ліва половина [0.5; 0.75], інтервал звужується: [0.5; 0.75]. Наступне число, **0**: інтервал [0.5; 0.75] ділиться навпіл — [0.5; 0.625] й [0.625; 0.75], — вибирається ліва половина [0.5; 0.625], інтервал звужується: [0.5; 0.625]. Наступне число, **1**: інтервал [0.5; 0.625] ділиться навпіл — [0.5; 0.5625] й [0.5625; 0.625], — вибирається права половина [0.5625; 0.6250], інтервал звужується: [0.5625; 0.6250].

За умовою точності задачі рішення знайдене: їм є будь-яке число з інтервалу [0.5625; 0.6250], наприклад, 0.625.

У принципі, якщо підходити строго, те ділення інтервалів потрібно продовжити доти, поки ліва й права границі знайденого інтервалу не ЗБІЖАТЬСЯ між собою з точністю до третього знака після коми. Тобто з позицій точності згенероване число вже не буде відмінно від будь-якого числа з інтервалу, у якому воно перебуває.

Другий спосіб рішення задачі

Розіб'ємо отриману двійкову послідовність 100110100 на тріади: 100, 110, 100. Після перекладу цих двійкових чисел у десяткові одержуємо: 4, 6, 4. Підставивши попереду «0.», одержимо: 0.464. Таким методом можуть виходити тільки числа від 0.000 до 0.777 (тому що максимум, що можна «вичавити» із трьох двійкових розрядів — це $111_2 = 7_8$) — тобто, по суті, ці числа представлені у восьмеричній системі числення. Для перекладу *восьмеричного* числа в *десяткове* подання виконаємо:

$0.464_8 = 4 \cdot 8^{-1} + 6 \cdot 8^{-2} + 4 \cdot 8^{-3} = 0.6015625_{10} = 0.602_{10}$.

Отже, шукане число дорівнює: 0.602.

Табличні ГВЧ

Табличні ГВЧ як джерело випадкових чисел використають спеціальним образом складені таблиці, що містять перевірені некорельовані, тобто ніяк не залежний друг від друга, цифри. У табл. 22.1 наведений невеликий фрагмент такої таблиці. Обходячи таблицю ліворуч праворуч зверху вниз, можна одержувати рівномірно розподілені від 0 до 1 випадкові числа з потрібним числом знаків після коми (у нашому прикладі ми використаємо для кожного числа по трьох знака). Тому що цифри в таблиці не залежать друг від друга, то таблицю можна обходити різними способами, наприклад, зверху вниз, або праворуч ліворуч, або, скажемо, можна вибирати цифри, що перебувають на парних позиціях.

Таблиця 22.1.
Випадкові цифри. Рівномірно розподілені від 0 до 1 випадкові числа

Випадкові цифри								**Рівномірно розподілені від 0 до 1 випадкові числа**
9	2	9	2	0	4	2	6	0. 929
9	5	7	3	4	9	0	3	0. 204
5	9	1	6	6	5	7	6	0. 269
…								…

Достоїнство даного методу в тім, що він дає дійсно випадкові числа, тому що таблиця містить перевірені некорельовані цифри. Недоліки методу: для зберігання великої кількості цифр потрібно багато пам'яті; більші труднощі породження й перевірки такого роду таблиць, повтори при використанні таблиці вже не гарантують випадковості числової послідовності, а виходить, і надійності результату.

Алгоритмічні ГВЧ

Числа, генеровані за допомогою цих ГВЧ, завжди є псевдовипадковими (або квазівипадковими), тобто кожне наступне згенероване число залежить від попередні:

$$r_{i+1} = f(r_i).$$

Послідовності, складені з таких чисел, утворять петлі, тобто обов'язково існує цикл, що повторюється нескінченне число раз. Повторювані цикли називаються періодами.

Достоїнством даних ГВЧ є швидкодія; генератори практично не вимагають ресурсів пам'яті, компактні. Недоліки: числа не можна повною мірою назвати випадковими, оскільки між ними є залежність, а також наявність періодів у послідовності квазислучайных чисел.

Розглянемо кілька алгоритмічних методів одержання ГВЧ:

- метод серединних квадратів;
- метод серединних добутків;
- метод перемішування;
- лінійний конгруентний метод.

Метод серединних квадратів

Є деяке чотиризначне число $R0$. Це число зводиться у квадрат і заноситься в $R1$. Далі з $R1$ береться середина (чотири середніх цифри) — нове випадкове число — і записується в $R0$. Потім процедура повторюється (див. мал. 22.6). Відзначимо, що насправді як випадкове число необхідно брати не **ghij**, а **0.ghij** — із приписаним ліворуч нулем і десятковою крапкою. Цей факт відбитий як на мал. 22.6, так і на наступних подібних малюнках.

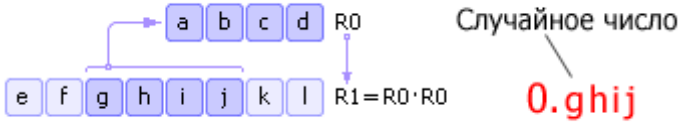


Рис. 22.6. Схема методу серединних квадратів

Недоліки методу: 1) якщо на деякій ітерації число $R0$ стане рівним нулю, то генератор вироджується, тому важливо правильний вибір початкового значення $R0$; 2) генератор буде повторювати послідовність через M^n кроків (у найкращому разі), де n — розрядність числа $R0$, M — підстава системи числення.

Для приклада на мал. 22.6: якщо число $R0$ буде представлено у двійковій системі числення, то послідовність псевдовипадкових чисел повториться через $2^4 = 16$ кроків. Помітимо, що повторення послідовності може відбутися й раніше, якщо початкове число буде обрано невдало.

Описаний вище спосіб був запропонований Джоном фон Нейманом і ставиться до 1946 року. Оскільки цей спосіб виявився ненадійним, від нього дуже швидко відмовилися.

Метод серединних добутків

Число $R0$ множиться на $R1$, з отриманого результату $R2$ витягається середина $R2^*$ (це чергове випадкове число) і множиться на $R1$. За цією схемою обчислюються всі наступні випадкові числа (див. мал. 22.7).

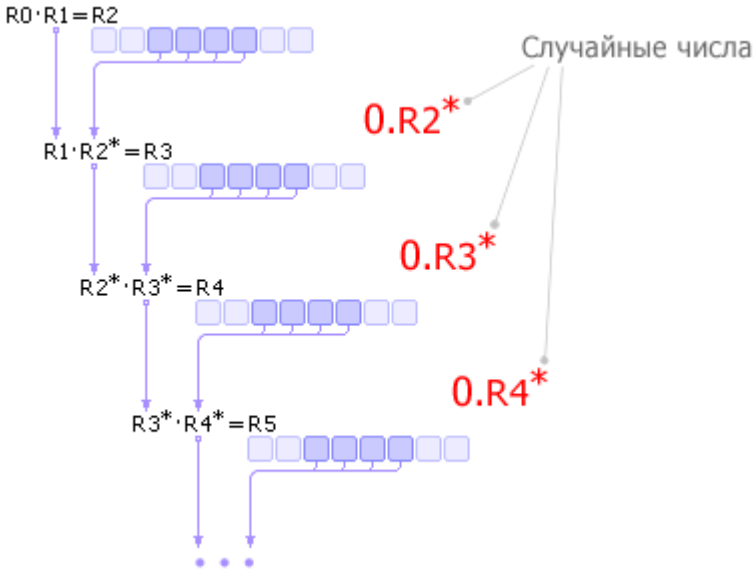


Рис. 22.7. Схема методу серединних добутків

Метод перемішування

У методі перемішування використаються операції циклічного зрушення вмісту осередку вліво й вправо. Ідея методу полягає в наступному. Нехай в осередку зберігається початкове число $R0$. Циклічно зрушуючи вміст осередку вліво на 1/4 довжини осередку, одержуємо нове число $R0^{*}$. Точно так само, циклічно зрушуючи вміст осередку $R0$ вправо на 1/4 довжини осередку, одержуємо друге число $R0^{**}$. Сума чисел $R0^{*}$ й $R0^{**}$ дає нове випадкове число $R1$. Далі $R1$ заноситься в $R0$, і вся послідовність операцій повторюється (див. мал. 22.8).

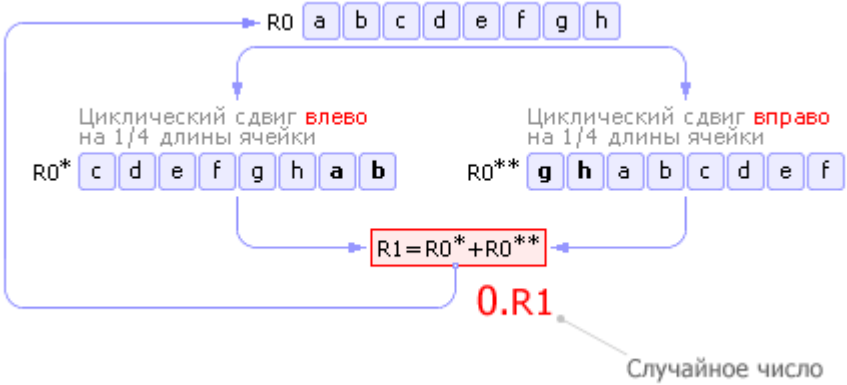


Рис. 22.8. Схема методу перемішування

Зверніть увагу, що число, отримане в результаті підсумовування $R0^*$ й $R0^{**}$, може не вміститися повністю в осередку $R1$. У цьому випадку від отриманого числа повинні бути відкинуті зайві розряди. Пояснимо це для мал. 22.8, де всі осередки представлені вісьма двійковими розрядами. Нехай $R0^* = 10010001_2 = 145_{10}$, $R0^{**} = 10100001_2 = 161_{10}$, тоді $R0^* + R0^{**} = 100110010_2 = 306_{10}$. Як бачимо, число 306 займає 9 розрядів (у двійковій системі числення), а осередок $R1$ (як й $R0$) може вмістити в себе максимум 8 розрядів. Тому перед занесенням значення в $R1$ необхідно забрати один «зайвий», крайній лівий біт із числа 306, у результаті чого в $R1$ піде вже не 306, а $00110010_2 = 50_{10}$. Також помітимо, що в таких мовах, як Паскаль, «урізування» зайвих бітів при переповненні осередку виробляється автоматично відповідно до заданого типу змінної.

Лінійний конгруентний метод

Лінійний конгруентний метод є однієї з найпростіших і найбільш уживаних у цей час процедур, що імітують випадкові числа. У цьому методі використається операція $\mathrm{mod}(x, y)$, що повертає остачу від ділення першого аргументу на другий. Кожне наступне випадкове число розраховується на основі попереднього випадкового числа по наступній формулі:

$$r_{i+1} = \mathrm{mod}(k \cdot r_i + b, M).$$

M — модуль ($0 < M$);

k — множник ($0 \leq k < M$);

b — приріст ($0 \leq b < M$);

r_0 — початкове значення ($0 \leq r_0 < M$).

Послідовність випадкових чисел, отриманих за допомогою даної формули, називається лінійною конгруентною послідовністю. Багато авторів називають лінійну конгруентну послідовність при $b = 0$ мультиплікативним конгруентним методом, а при $b \neq 0$ — змішаним конгруентним методом.

Для якісного генератора потрібно підібрати підходящі коефіцієнти. Необхідно, щоб число M було досить більшим, тому що період не може мати більше M елементів. З іншого боку, ділення, що використається в цьому методі, є досить повільною операцією, тому для двійкової обчислювальної машини логічним буде вибір $M = 2^N$, оскільки в цьому випадку знаходження остачі від ділення зводиться усередині ЕОМ до двійкової логічної операції «AND». Також широко розповсюджений вибір найбільшого простого числа M, меншого, чим 2^N: у спеціальній літературі доводиться, що в цьому випадку молодші розряди одержуваного випадкового числа r_{i+1} поводяться так само випадково, як і старші, що позитивно позначається на всій послідовності випадкових чисел у цілому. Як приклад можна привести одне із *чисел Мерсена*, рівне $2^{31} - 1$, і таким чином, $M = 2^{31} - 1$.

Однією з вимог до лінійних конгруентних послідовностей є як можна більша довжина періоду. Довжина періоду залежить від значень M, k й b. Теорема, що ми приведемо нижче, дозволяє визначити, чи можливо досягнення періоду максимальної довжини для конкретних значень M, k й b.

Теорема. Лінійна конгруентна послідовність, певна числами M, k, b й r_0, має період довжиною M тоді й тільки тоді, коли:

- числа b й M взаємно прості;
- $k - 1$ кратно p для кожного простого p, що є дільником M;
- $k - 1$ кратно 4, якщо M кратно 4.

Нарешті, на закінчення розглянемо пари прикладів використання лінійного конгруентного методу для генерації випадкових чисел.

Приклад 1

$M = 2^N$
$k = 3 + 8 \cdot q$ (або $k = 5 + 8 \cdot q$)
$b = 0$
r_0 — непарне

Було встановлено, що ряд псевдовипадкових чисел, генерованих на основі даних із приклада 1, буде повторюватися через кожні $M/4$ чисел. Число q задається довільно перед початком обчислень, однак при цьому варто мати на увазі, що ряд робить враження випадкового при більших k (а виходить, і q). Результат можна трохи поліпшити, якщо b непарне й $k = 1 + 4 \cdot q$ — у цьому випадку ряд буде повторюватися через кожні M чисел. Після довгих пошуків k дослідники зупинилися на значеннях 69069 й 71365.

Приклад 2

$M = 2^{31} - 1$
$k = 1\ 220\ 703$
125
$b = 7$
$r_0 = 7$

Генератор випадкових чисел, що використає дані із приклада 2, буде видавати випадкові неповторювані числа з періодом, рівним 7 мільйонам.

Мультиплікативний метод генерації псевдовипадкових чисел був запропонований Д. Г. Лехмером (D. H. Lehmer) в 1949 році.

Перевірка якості роботи генератора

Від якості роботи ГВЧ залежить якість роботи всієї системи й точність результатів. Тому випадкова послідовність, породжувана ГВЧ, повинна задовольняти цілому ряду критеріїв.

Здійснювані перевірки бувають двох типів:

- перевірки на рівномірність розподілу;
- перевірки на статистичну незалежність.

Перевірки на рівномірність розподілу

1) ГВЧ повинен видавати близькі до наступного значення статистичних параметрів, характерних для рівномірного випадкового закону:

$$m_r = \frac{\sum_{i=1}^{n} r_i}{n} \approx 0.5$$ — математичне очікування;

$$D_r = \frac{\sum_{i=1}^{n} (r_i - m_r)^2}{n} \approx 0.0833$$ — дисперсія;

$$\sigma_r = \sqrt{D_r} \approx 0.2887$$ — середньоквадратичне відхилення.

2) Частотний тест

Частотний тест дозволяє з'ясувати, скільки чисел потрапило в інтервал $(m_r - \sigma_r; m_r + \sigma_r)$, тобто $(0.5 - 0.2887; 0.5 + 0.2887)$ або, в остаточному підсумку, $(0.2113; 0.7887)$. Тому що $0.7887 - 0.2113 = 0.5774$, містимо, що в гарному ГВЧ у цей інтервал повинне попадати близько 57.7% із всіх випадкових чисел, що випали, (див. мал. 22.9).

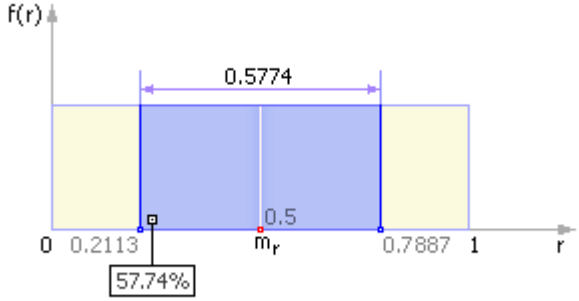


Рис. 22.9. Частотна діаграма ідеального ГВЧ у випадку перевірки його на частотний тест

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

3) Перевірка за критерієм «хі-квадрат»

Критерій «хі-квадрат» ($\chi^{2\text{-критерій}}$) — це один з найвідоміших статистичних критеріїв; він є основним методом, використовуваним у сполученні з іншими критеріями. Критерій «хі-квадрат» був запропонований в 1900 році Карлом Пірсоном. Його чудова робота розглядається як фундамент сучасної математичної статистики.

Для нашого випадку перевірка за критерієм «хі-квадрат» дозволить довідатися, наскільки створений нами *реальний* ГВЧ близький до еталона ГВЧ, тобто чи задовольняє він вимозі рівномірного чи розподілу ні.

Частотна діаграма *еталонного* ГВЧ представлена на мал. 22.10. Тому що закон розподілу еталонного ГВЧ рівномірний, те (теоретична) імовірність p_i влучення чисел в *i-ий* інтервал (усього цих інтервалів k) дорівнює $p_i = 1/k$. І, таким чином, у кожний з k інтервалів потрапить *рівно* по $p_i \cdot N$ чисел (N — загальна кількість згенерованих чисел).

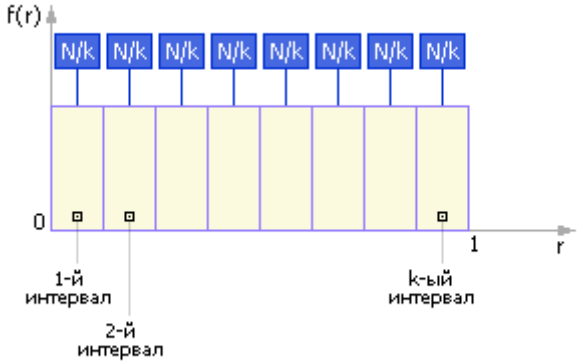


Рис. 22.10. Частотна діаграма еталонного ГВЧ

Реальний ГВЧ буде видавати числа, розподілені (причому, не обов'язково рівномірно!) по k інтервалах й у кожен інтервал потрапить по n_i чисел (у сумі $n_1 + n_2 + \ldots + n_k = N$)... Як же нам визначити, наскільки випробовуваний ГВЧ гарний і близький до еталонного? Цілком логічно розглянути квадрати різниць між отриманою кількістю чисел n_i й «еталонним» $p_i \cdot N$. Складемо їх, і в результаті одержимо:

$$\chi^2_{\text{експ.}} = (n_1 - p_1 \cdot N)^2 + (n_2 - p_2 \cdot N)^2 + \ldots + (n_k - p_k \cdot N)^2 \ldots$$

Із цієї формули треба, що чим менше різниця в кожному із що складають (а виходить, і чим менше значення $\chi^2_{\text{експ.}}$), тим сильніше закон розподілу випадкових чисел, генерованих реальним ГВЧ, тяжіє до рівномірного.

У попереднім вираженні кожному з доданків приписується однакова вага (рівний 1), що насправді може не відповідати дійсності; тому для статистики «хі-квадрат» необхідно провести нормування кожного *і-го* доданка, поділивши його на $p_i \cdot N$:

$$\chi^2_{\text{эксп.}} = \frac{(n_1 - p_1 \cdot N)^2}{p_1 \cdot N} + \frac{(n_2 - p_2 \cdot N)^2}{p_2 \cdot N} + \ldots + \frac{(n_k - p_k \cdot N)^2}{p_k \cdot N}$$

Нарешті, запишемо отримане вираження більш компактно й спростимо його:

Ми одержали значення критерію «хі-квадрат» для *експериментальних* даних.

У табл. 22.2 наведені *теоретичні* значення «хі-квадрат» ($\chi^2_{\text{теор.}}$), де $\nu = N - 1$ — це число ступенів волі, **p** — це довірча ймовірність, задана користувачем, що вказує, наскільки ГВЧ повинен задовольняти вимогам рівномірного розподілу, або **p** — *це ймовірність того, що експериментальне значення* $\chi^2_{\text{експ.}}$ *буде менше табуьованого (теоретичного)* $\chi^2_{\text{теор.}}$ *або дорівнює йому*.

Таблиця 22.2.

Деякі процентні крапки 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

$\boldsymbol{v}$ **= 11**	3. 053	4. 575	7. 584	10.34	13.70	19.68	24.72
$\boldsymbol{v}$ **= 12**	3. 571	5. 226	8. 438	11.34	14.85	21.03	26.22
$\boldsymbol{v}$ **= 15**	5. 229	7. 261	11.04	14.34	18.25	25.00	30.58
$\boldsymbol{v}$ **= 20**	8. 260	10.85	15.45	19.34	23.83	31.41	37.57
$\boldsymbol{v}$ **= 30**	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$\boldsymbol{v}$ **= 50**	29.71	34.76	42.94	49.33	56.33	67.50	76.15
$\boldsymbol{v}$ **> 30**	$v + \mathrm{sqrt}(2v) \cdot x_p + 2/3 \cdot x^2{}_p - 2/3 + O(1/\mathrm{sqrt}(v))$						
$\boldsymbol{x_p}$ **=**	–2.33	–1.64	–0. 674	0.00	0. 674	1.64	2.33

Прийнятним уважають **p** *від 10% до 90%.*

Якщо $\chi^2_{\text{експ.}}$ багато більше $\chi^2_{\text{теор.}}$ (тобто **p** — велике), те генератор **не задовольняє** вимозі рівномірного розподілу, тому що спостережувані значення n_i занадто далеко йдуть від теоретичних $p_i \cdot N$ і не можуть розглядатися як випадкові. Інакше кажучи, установлюється такий великий довірчий інтервал, що обмеження на числа стають дуже нежорсткими, вимоги до чисел - слабкими. При цьому буде спостерігатися дуже більша абсолютна погрішність.

Ще Д. Батіг (більш відомий як Дональд Кнут) у своїй книзі «Мистецтво програмування» помітив, що мати $\chi^2_{\text{експ.}}$ маленьким теж, взагалі ж, недобре, хоча це й здається, на перший погляд, чудово з погляду рівномірності. Дійсно, візьміть ряд чисел 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, … — вони ідеальні з погляду рівномірності, і $\chi^2_{\text{експ.}}$ буде практично нульовим, але навряд чи ви їх визнаєте випадковими.

Если $\chi^2_{\text{експ.}}$ багато менше $\chi^2_{\text{теор.}}$ (тобто **p** — мало), те генератор **не задовольняє** вимозі випадкового рівномірного розподілу, тому що спостережувані значення n_i занадто близькі до теоретичних $p_i \cdot N$ і не можуть розглядатися як випадкові.

А от якщо $\chi^2_{\text{експ.}}$ лежить у деякому діапазоні, між двома значеннями $\chi^2_{\text{теор.}}$, які відповідають, наприклад, **p** = 25% й **p** = 50%, те можна вважати, що значення випадкових чисел, породжувані датчиком, цілком є випадковими.

При цьому додатково треба мати на увазі, що всі значення $p_i \cdot N$ повинні бути досить більшими, наприклад більше 5 (з'ясовано емпіричним шляхом). Тільки тоді (при досить великій статистичній вибірці) умови проведення експерименту можна вважати задовільними.

Отже, процедура перевірки має такий вигляд.

1. Діапазон від 0 до 1 розбивається на k рівних інтервалів.
2. Запускається ГВЧ N раз (N повинне бути велике, наприклад, $N/k > 5$).
3. Визначається кількість випадкових чисел, що потрапили в кожен інтервал: n_i, $i = 1, \ldots, k$...
4. Обчислюється експериментальне значення $\chi^2_{\text{експ.}}$ по наступній формулі:

$$\chi^2_{\text{эксп.}} = \sum_{i=1}^{k} \frac{(n_i - p_i \cdot N)^2}{p_i \cdot N} = \frac{1}{N} \sum_{i=1}^{k} \left(\frac{n_i^2}{p_i} \right) - N$$

де $p_i = 1/k$ — теоретична ймовірність влучення чисел в k-*ий* інтервал.

5. Шляхом порівняння експериментально отриманого значення $\chi^2_{\text{експ.}}$ з теоретичним $\chi^2_{\text{теор.}}$ (з табл. 22.2) робиться висновок про придатність генератора для використання. Для цього: а) входимо в табл. 22.2 (**рядок = кількість експериментів – 1**); б) порівнюємо обчислене $\chi^2_{\text{експ.}}$ з $\chi^2_{\text{теор.}}$, що зустрічаються в рядку. При цьому можливо три випадки.

   Перший випадок: $\chi^2_{\text{експ.}}$ багато більше будь-якого $\chi^2_{\text{теор.}}$ у рядку — гіпотеза про випадковості рівномірного генератора не виконується (розкид чисел занадто великий, щоб бути випадковим).

   Другий випадок: $\chi^2_{\text{експ.}}$ багато менше будь-якого $\chi^2_{\text{теор.}}$ у рядку — гіпотеза про випадковості рівномірного генератора не виконується (розкид чисел занадто малий, щоб бути випадковим).

Третій випадок: $\chi^2_{\text{експ.}}$ лежить між значеннями $\chi^2_{\text{теор.}}$ двох рядом вартих стовпців — гіпотеза про випадковості рівномірного генератора виконується з імовірністю **p** (тобто в **p** випадках з 100).

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

Перевірки на статистичну незалежність

1) Перевірка на частоту появи цифри в послідовності

Розглянемо приклад. Випадкове число 0.2463389991 складається із цифр 2463389991, а число 0.5467766618 складається із цифр 5467766618. З'єднуючи послідовності цифр, маємо: 24633899915467766618.

Зрозуміло, що теоретична ймовірність p_i випадання *і-оїї* цифри (від 0 до 9) дорівнює 0.1.

Далі варто обчислити частоту появи кожної цифри в експериментальній послідовності, що випала. Наприклад, цифра 1 випала 2 рази з 20, а цифра 6 випала 5 разів з 20.

Далі вважають оцінку й ухвалюють рішення щодо критерію «хі-квадрат».

2) Перевірка появи серій з однакових цифр

Позначимо через n число серій однакових підряд цифр довжини L. Перевіряти треба всі L від 1 до m, де m — це задане користувачем число: максимально, що зустрічається число, однакових цифр у серії.

У прикладі «24633899915467766618» виявлені 2 серії довжиною в 2 (33 й 77), тобто $n_2 = 2$ й 2 серії довжиною в 3 (999 й 666), тобто $n_3 = 2$.

Вероятность появи серії довжиною в L дорівнює: $p = 9 \cdot 10^{-L}$ (теоретична). Тобто ймовірність появи серії довжиною в один символ дорівнює: $p_1 = 0.9$ (теоретична). Імовірність появи серії довжиною у два символи дорівнює: $p_2 = 0.09$ (теоретична). Імовірність появи серії довжиною в три символи дорівнює: $p_3 = 0.009$ (теоретична).

Наприклад, імовірність появи серії довжиною в один символ дорівнює $p = 0.9$, тому що всього може зустрітися один символ з 10, а всього символів 9 (нуль не вважається). А ймовірність того, що підряд зустрінеться два однакових символи «XX» дорівнює $0.1 \cdot 0.1 \cdot 9$, тобто ймовірність 0.1 того, що в першій позиції з'явиться символ «X», множиться на ймовірність 0.1 того, що в другій позиції з'явиться такий же символ «X» і множиться на кількість таких комбінацій 9.

Частость появи серій підраховується по раніше розібраній нами формулі «хі-квадрат» з використанням значень p.

Примітка: генератор може бути перевірений багаторазово, однак перевірки не мають властивість повноти й не гарантують, що генератор видає випадкові числа. Наприклад, генератор, що видає послідовність 12345678912345..., при перевірках буде вважатися ідеальним, що, мабуть, не зовсім так.

На закінчення відзначимо, що третій розділ книги Дональда Е. Кнута «Мистецтво програмування» (тім 2) повністю присвячена вивченню випадкових чисел. У ній вивчаються різні методи генерування випадкових чисел, статистичні критерії випадковості, а також перетворення рівномірно розподілених випадкових чисел в інші типи випадкових величин. Викладу цього матеріалу приділено більше двохсот сторінок.