МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»
(Самарский университет)

И.С. Орлова

Методы защиты информации, использующие генераторы псевдослучайных чисел

Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве учебного пособия для студентов, обучающихся по программе высшего образования по направлениям подготовки бакалавров 10.03.01 Информационная безопасность, 03.03.01 Прикладная математика и физика, 02.03.02 Фундаментальная информатика и информационные технологии

С А М А Р А
Издательство Самарского университета
2016

УДК 519.6(075)+004.9(075)
ББК 221я7+32.81я7
О-664

Рецензенты: д-р физ.-мат. наук, проф. С. Я. Шатских,
к-т физ.-мат. наук, доц. А. Ю. Трусова.

О-664 ***Орлова, Ирина Сергеевна***
Методы защиты информации, использующие генераторы псевдослучайных чисел: учеб. пособие / *И.С. Орлова*. – Самара: Изд-во Самарского университета, 2016. – 48 с.

ISBN 978-5-7883-1089-3

В данном пособии представлены основные виды генераторов псевдослучайных чисел, создающих (с помощью математических алгоритмов, реализуемых на компьютерах) неслучайные числовые последовательности, свойства которых близки к «типичным» свойствам последовательностей независимых равномерно распределенных случайных величин. Рассматриваются вопросы оценки качества (статистическое тестирование) генераторов псевдослучайных чисел, а также их применение в задачах защиты информации, а также в методах Монте-Карло.

Предназначено для студентов специальностей и направлений 10.03.01 Информационная безопасность, 03.03.01 Прикладная математика и физика, 02.03.02 Фундаментальная информатика и информационные технологии.

УДК 519.6(075)+004.9(075)

ISBN 978-5-7883-1089-3 ББК 221я7+32.81я7

Оглавление

Введение

Последовательности случайных и псевдослучайных чисел

$$x_1, \ldots, x_m, \ldots$$

обладают свойствами, близкими к «типичным» свойствам последовательности числовых значений независимых случайных величин

$$X_1, \ldots, X_m, \ldots$$

с общей функцией распределения

$$F(x) = \mathbb{P}\{X < x\}.$$

Согласно принятой в настоящее время терминологии, различают последовательности *случайных чисел*, производимые генераторами случайных чисел (ГСЧ), которые используют физические хаотические процессы (радиоактивный распад, фотоэлектрический эффект, шумы полупроводниковых диодов, космическое излучение и др.), и последовательности *псевдослучайных чисел*, которые генерируются с помощью математических алгоритмов, реализуемых на компьютерах (генераторы псевдослучайных чисел (ГПСЧ)).

К упомянутым типичным свойствам последовательностей случайных и псевдослучайных чисел относят: а) устойчивость частот появления числовых значений, б) непредсказуемость появления того или иного числового значения, в) сложную структуру (компьютерные программы для генерации таких последовательностей должны иметь большую длину) и т.д.

В настоящее время не существует методов построения идеальных числовых последовательностей, которые бы обладали полным набором свойств последовательностей числовых значений независимых одинаково распределенных случайных величин[1].

[1] См. [1], глава 3.

Заметим, что в рамках теории вероятностей понятие (индивидуальной) числовой случайной последовательности не рассматривается.

Генераторы случайных и псевдослучайных чисел имеют широкую область применения: статистическое моделирование (метод Монте-Карло), защита информации (включая криптографические методы защиты), различные приложения математической статистики и т.д. При этом от качества используемых ГСЧ и ГПСЧ напрямую зависит качество получаемых результатов.

В этом пособии будут рассмотрены основные виды ГПСЧ, а также методы оценки качества вырабатываемых ими псевдослучайных последовательностей.

Глава 1. Примеры генераторов псевдослучайных чисел

1.1. Линейный конгруэнтный генератор

Выберем натуральное число m (*модуль*) и три целых числа:

начальное значение $x_0 \in \mathbb{Z}_m \equiv \{0, 1, \ldots, m-1\}$;
множитель $a \in \mathbb{Z}_m$; *приращение* $c \in \mathbb{Z}_m$.

Определение. Последовательность целых чисел, получаемая с помощью соотношения

$$x_{n+1} = ax_n + c\,(mod\, m),\ n = 0, 1, 2, \ldots, \tag{1.1}$$

называется *линейной конгруэнтной последовательностью* (ЛКП). Само соотношение (1.1) называют *линейным конгруэнтным генератором* (ЛКГ).

Нетрудно видеть, что такая ЛКП периодична и её период не превышает m. Действительно, ввиду того, что для любого натурального n

$$x_n \in \mathbb{Z}_m,$$

то среди первых $m+1$ членов этой последовательности обязательно найдутся по крайней мере два одинаковых, а это, ввиду равенства (1.1), влечет за собой периодичность ЛКП с периодом, не превышающим m.

Поскольку длинный период необходим для псевдослучайных последовательностей, которые используются в качестве случайных, то представляет интерес подбор параметров $\{a, c, m\}$, при котором период ЛКП достигает максимального значения, равного модулю m.

Теорема. Линейная конгруэнтная последовательность (1.1), определенная числами $\{a, c, m\}$, имеет при любом начальном значении x_0 максимальный период m тогда и только тогда, когда

1) числа c и m взаимно просты, т.е. $(c, m) = 1$;

2) число $b = a - 1$ кратно p для каждого простого $p < m$, являющегося делителем числа m;

3) число b кратно 4, если m кратно 4.

Без доказательства[2].

Примеры ЛКГ, имеющих максимальный период[3]

1. $\{a = 106,\ c = 1283,\ m = 6075\}$.

$(1283, 6075) = 1,$ так как $6075 = 3^5 \cdot 5^2,\ b = a - 1 = 105 = 3 \cdot 5 \cdot 7.$

2. $\{a = 141,\ c = 28411,\ m = 134456\}$.

$(28411, 134456) = 1,\ m = 134456 = 4 \cdot 2 \cdot 7^5,\ b = a - 1 = 140 = 4 \cdot 5 \cdot 7,$

Программы разложения на множители

MAXIMA

ifactors(134456) Shift Enter $[[2, 3], [7, 5]]$ $134456 = 2^3 \cdot 7^5$	**ifactors(28411)** Shift Enter $[[28411, 1]]$ 28411 - простое число	**ifactors(140)** Shift Enter $[[2, 2], [5, 1], [7, 1]]$ $140 = 2^2 \cdot 5 \cdot 7$

MATHEMATICA

FactorInteger[140] Shift Enter {{2, 2}, {5, 1},{7,1} $140 = 2^2 \cdot 5 \cdot 7$	**FactorInteger[134456]** Shift Enter {{2,3},{7,5}} 134456=$4 \cdot 2 \cdot 7^5$	**FactorInteger[28411]** Shift Enter {{28411, 1}} 28411 - простое число

[2]Доказательство смотри в [1], с. 28.
[3]См. [8], с. 416-417.

Программы нахождения наибольшего общего делителя[4]

MAXIMA

load(gcdex)$
Shift Enter
igcdex(28411,134456);
Shift Enter
[48499, -10248,**1**], $48499 \cdot 28411 - 10248 \cdot 134456 = 1.$

MATHEMATICA

GCD[28411, 134456]
Shift Enter
1
Пример. $x_{n+1} = 9x_n + 13\,(mod\,64),\ \ x_0 = 0.$

MATHEMATICA

Mod[a, b] = a (mod b),

$\mathbf{Table}[\mathbf{expr}, \mathbf{i}, \mathbf{i}_{\max}]$– генерируются значения $expr,$ когда $i = \overline{1, i_{\max}}.$

Table[Mod[9 i + 13, 64], {i, 64}]

22, 31, 40, 49, 58, 3, 12, 21, 30, 39, 48, 57, 2, 11, 20, 29, 38, 47,
56, 1, 10, 19, 28, 37 46, 55, 0, 9, 18, 27, 36, 45, 54, 63, 8, 17,
26, 35, 44, 53, 62, 7, 16, 25, 34, 43, 52, 61 6, 15, 24, 33, 6, 15,
24, 33, 42, 51, 60, 5, 14, 23, 32, 41, 50, 59, 4, 13, 22, 31.

Примечание. Линейные конгруэнтные генераторы имеют «слабость»: при рассмотрении последовательности биграмм $(z_n^{(1)},\ z_n^{(2)})$, где

$$z_n^{(1)} = x_n,\ z_n^{(2)} = x_{n+1}$$

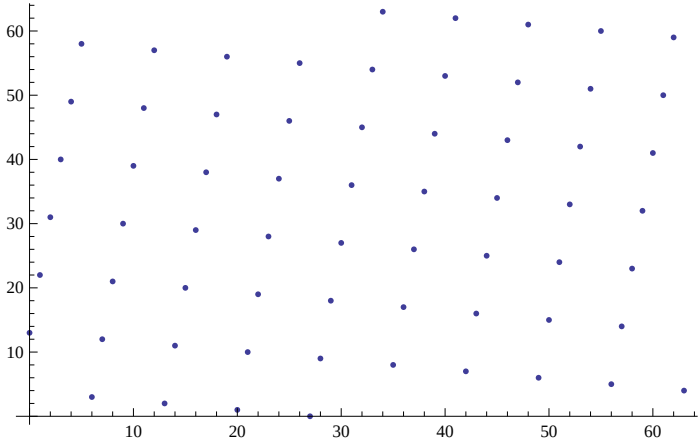
как точек на $(z^{(1)},\ z^{(2)})$ плоскости, эти точки будут лежать на прямых семейства

$$z^{(2)} = az^{(1)} + c - km,\ k = 0, 1, 2, \dots\ .$$

Пример. $x_{n+1} = 9x_n + 13\,(mod\,64),\ \ x_0 = 0.$

[4]**G**reatest **C**ommon **D**ivisor (англ.) – наибольший общий делитель.

$$z^{(2)} = 9z^{(1)} + 13,\ (k = 0);\ z^{(2)} = 9z^{(1)} - 51,\ (k = 1); \ldots$$



Примечание. Отмеченная «слабость» линейного конгруэнтного генератора позволяет «вскрывать» этот генератор, т.е. находить неизвестные параметры $\{a, c, m\}$ по нескольким наблюдаемым значениям x_n. Поэтому ЛКГ нельзя использовать в криптографии, но можно использовать в задачах статистического моделирования[5].

1.2. Квадратичный конгруэнтный генератор

Алгоритм генерации псевдослучайной последовательности задается квадратичным рекуррентным соотношением

$$x_{n+1} = dx_n^2 + ax_n + c\,(mod\, m), \tag{1.2}$$

где $x_0, a, c, d \in \mathbb{Z}_m$ и $x_n \in \mathbb{Z}_m,\ n = 1, 2, \ldots$.

Теорема. Квадратичная конгруэнтная последовательность, вырабатываемая квадратичным конгруэнтным генератором (1.2), имеет наибольший период m тогда и только тогда, когда выполнены следующие условия:

[5][8], с. 415, 417.

1) $(c, m) = 1$ т.е c и m – взаимно простые числа;

2) d и $a - 1$ – кратны q, где q – любой нечетный простой делитель модуля m;

3) d – четное число, причем

$$d = \begin{cases} (a-1)\, mod\, 4, & \text{если } m \text{ кратно } 4, \\ (a-1)\, mod\, 2, & \text{если } m \text{ кратно } 2; \end{cases}$$

4) если модуль m кратен 9, то $d \neq 3c\, mod\, 9$.

Без доказательства[6].

Пример. Рассмотрим квадратичный конгруэнтный генератор (ККГ) с параметрами $m = 36 = 2^2 \cdot 3^2,\ d = 12 = 2^2 \cdot 3,\ a = 25,\ c = 11$.

Проверим условия теоремы о максимальном периоде ККГ.

1) $(c, m) = (11, 36) = 1$.

2) $d = 12$ и $a - 1 = 24$ кратны 2 и 3 (нечетным простым делителям модуля $m = 36 = 2^2 \cdot 3^2$).

3)

$$12 = \begin{cases} 24\, mod\, 4, & \text{т.к. } m = 36 \text{ кратно } 4, \\ 24\, mod\, 2, & \text{т.к. } m = 36 \text{ кратно } 2. \end{cases}$$

4) Так как модуль $m = 36$ кратен 9, то предположение $12 = 3 \cdot 11\, mod\, 9$ приводит к противоречию $21 = 0\, mod\, 9$.

Для $x_0 = 1$, используя равенство (1.2), получим множество всех целых чисел от 0 до 35.

MATHEMATICA

$$\mathbf{Table[Mod[12i^2 + 25i + 11, 36], \{i, 36\}]}$$

[6] См. [1], с. 49.

12, 23, 22, 33, 8, 7, 18, 29, 28, 3, 14, 13, 24, 35, 34, 9, 20, 19,
30, 5, 4, 15, 26, 25 0, 11, 10, 21, 32, 31, 6, 17, 16, 27, 2, 1.

Примечание. Существует алгоритм Джоан Бояр (J. Boyar), позволяющий вскрывать квадратические конгруэнтные генераторы,[7] т.е. находить неизвестные параметры $\{d, a, c, m\}$ по нескольким наблюдаемым значениям x_n.

1.3. Линейный регистр сдвига с обратной связью (LFSR[8])

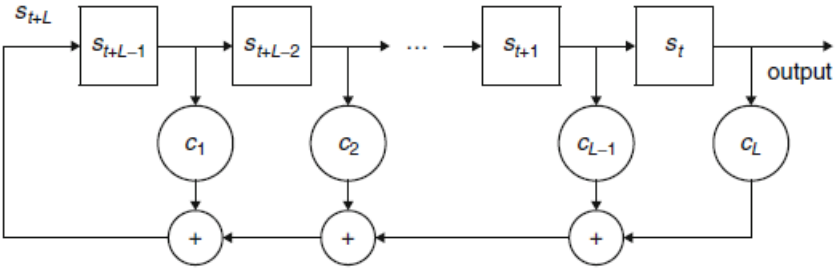
Линейный регистр сдвига с обратной связью генерирует последовательность чисел из конечного поля $\mathbb{F}_q = \{0, 1, \ldots, q-1\}$,

$$\mathbf{s} = \{s_0, s_1, \ldots, s_m, \ldots\},$$

которые удовлетворяют линейному разностному уравнению степени L

$$s_{t+L} = \sum_{i=1}^{L} c_i \, s_{t+L-i} \, (mod\, q), \; t = 1, 2, \ldots \tag{1.3}$$

с *заданными коэффициентами* $\mathrm{c}_i \in \mathbb{F}_q, \; i = \overline{1, L}$.



LFSR реализуется с помощью L ячеек памяти, состояние которых в момент времени $t = 0, 1, 2, \ldots$ характеризуется значениями

$$s_t, s_{t+1}, \ldots, s_{t+L-1} \in \mathbb{F}_q.$$

[7] См. [8], с. 415.
[8] Linear Feedback Shift Register.

Выходы ячеек памяти связаны последовательно друг с другом, а также с сумматорами с коэффициентами c_i.

В начальный момент времени $t = 0$ регистр загружается *произвольно выбранными числами (начальное состояние регистра)*

$$s_0, s_1, \ldots, s_{L-1} \in \mathbb{F}_q \tag{1.4}$$

и по формуле (1.3) генерирует число

$$s_L = \mathrm{c}_1 s_{L-1} + \ldots + \mathrm{c}_L s_0 \,(mod\, q).$$

В следующий момент времени $t = 1$ регистр генерирует число

$$s_{L+1} = \mathrm{c}_1 s_L + \ldots + \mathrm{c}_L s_1 \,(mod\, q).$$

Наконец, в момент времени $t > 0$ по формуле (1.3) регистр генерирует число

$$s_{L+t} = \mathrm{c}_1 s_{L+t-1} + \ldots + \mathrm{c}_L s_t \,(mod\, q). \tag{1.5}$$

Таким образом, на основе набора коэффициентов $\{c_1, \ldots, c_L\}$ и начального состояния $\{s_0, \ldots, s_{L-1}\}$ LFSR генерирует псевдослучайную числовую последовательность

$$s_L, s_{L+1}, \ldots, s_{L+t}, \text{ в поле } \mathbb{F}_q.$$

Периодичность LFSR. Начальными состояниями регистра сдвига (1.3) (при фиксированных коэффициентах c_i) полностью определяется генерируемая последовательность. Так как число всех возможных ненулевых начальных состояний регистра равно $q^L - 1$, то генерируемая последовательность периодична с периодом $T \leq q^L - 1$. Условия выбора коэффициентов c_i, при которых достигается максимальная длина периода $T_{\max} = q^L - 1$, приведены в [2] на стр. 198, 199.

Пример. Построить псевдослучайную числовую последовательность, если генератор LFSR обладает следующими параметрами: $q = 2,\ L = 4,$

$$\{c_1, c_2, c_3, c_4\} = \{0, 0, 1,\ 1\},\ \ \{s_0,\ s_1, s_2, s_3\} = \{1, 0, 1, 1\}.$$

Решение. Из формулы (1.5) следует равенство

$$s_{4+t} = 0 \cdot s_{t+3} + 0 \cdot s_{t+2} + 1 \cdot s_{t+1} + 1 \cdot s_t \, (mod\, 2).$$

MATHEMATICA

Программа $\mathbf{LinearRecurrence}[\{\mathbf{c_1}, \ldots, \mathbf{c_d}\}, \{\mathbf{s_1}, \ldots, \mathbf{s_d}\}, \mathbf{n}]$

генерирует числовую последовательность $\{s_k\}_{k=1}^{n}$, удовлетворяющую рекуррентному соотношению

$$s_{k+d} = c_1 s_{k+d-1} + \ldots + c_d s_k,$$

с начальными условиями $s_1, \ldots, s_d$.

Для вычисления значений членов предыдущей числовой последовательности по $mod\, 2$ используют программу

$$\mathbf{Mod}[\mathbf{LinearRecurrence}[\{\mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{1}\}, \{\mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{1}\}, \mathbf{19}], \mathbf{2}]$$

$$\underbrace{\mathbf{1,0,1,1,1,1,0,0,0,1,0,0,1,1,0}}_{15}, 1, 0, 1, 1; \; T_{\max} = 2^4 - 1 = 15.$$

Примечание. Вскрытие LFSR-генератора псевдослучайных чисел можно провести с помощью алгоритма Берлекэмпа-Месси[9] .

1.4. BBS-генератор псевдослучайных последовательностей[10]

Алгоритм генерации двоичной псевдослучайной последовательности

$$\{x_1, x_2, \ldots, x_n\}, \; x_k \in \mathbb{F}_2 = \{0, 1\}.$$

1. Генерируются два достаточно больших *секретных различных простых* числа p и q, таких, что

$$p \; = \; 3 \, (mod\, 4), \quad q = 3 \, (mod\, 4).$$

[9] [2], с. 219.

[10] Blum L., Blum M., Shub M. (1986 г.).

2. Вычисляется число $m = p \cdot q$.

3. Выбирается случайное *стартовое* натуральное число

$$s \in \{1, 2, \ldots, m-1\} \text{ такое, что } (s,\, m) = 1.$$

4. Вычисляется число $u_0 = s^2 \,(mod\, m)$.

5. Вычисляются числа $u_{k+1} = u_k^2 \,(mod\, m),\ k = 0, 1, 2, \ldots$.

6. Вычисляется x_k как самый младший бит двоичного представления числа $x_k = u_k \,(mod\, 2)$.

7. Формируется последовательность

$$\{x_1, x_2, \ldots, x_n\},\ , x_k \in \mathbb{F}_2 = \{0, 1\}.$$

Примечание. Ввиду равенства

$$u_k = u_0^{2^k (mod (p-1)(q-1))} \,(mod\, m), \quad k = 1, 2, \ldots,$$

для вычисления x_k необязательно иметь в распоряжении все $k-1$ предыдущих значений x_n.

BBS-генератор ПСЧ используется в криптографии как достаточно стойкий ко взлому генератор. Стойкость генератора обеспечивается вычислительной сложностью задачи разложения натурального числа m на множители. Использование BBS-генератора ПСЧ в задачах моделирования затруднена недостаточно высокой скоростью работы этого генератора.

Пример 1. $p = 7,\ (p - 3 = 4),\ q = 11,\ (q - 3 = 8),\ m = 77,$ $(p-1)(q-1) = 60,$ выберем $s = 30$, для которого $(77, 30) = 1$.

$$u_0 = 30^2 \,(mod\, 77) = 53,\ u_1 = 53^2 \,(mod\, 77) = 37,$$

$$u_2 = 37^2 \,(mod\, 77) = 60,\ u_3 = 60^2 \,(mod\, 77) = 58,$$

$$u_4 = 58^2 \,(mod\, 77) = 53,\ \textbf{период!}$$

В результате получаем (0,1) - последовательность

$$x_1 = 1,\ x_2 = 0,\ x_3 = 0,\ x_4 = 1, \quad \textbf{период!}$$

MATHEMATICA

$$\mathbf{Mod[Table\left[Mod\left[30^{Mod[2^k,60]},77\right],\{k,2,9\}\right],2]}$$

1, 0, 0, 1, 1, 0, 0, 1 период!

Пример 2. p = 67 = 3 (mod 4), q = 131 = 3 (mod 4),
p q = 8777, (p-1)(q-1) = 8530,
выберем $s = 11$, для которого $(8777, 30) = 1$.

MATHEMATICA

$$\mathbf{Table\left[Mod\left[11^{Mod[2^k,8530]},8777\right],\{k,2,62\}\right]}$$

5864,6987,495,8046,7741,2502,2003,920,3808,1260,7740,4575,6257,4629,2984,
4378,6693,7218,8029,6553,4725,5714,8133,2217,8746,961,1936,317,3942,4074,
169,2230,5118,3356,1845,7326,7698,5677,7962,5950,4859,8528,562,8649,7607,
8465,797,3265,4947,2533,102,1627,5252,6170,3051,4981,6559,4404,6823,121,

5864.

$$\mathbf{Mod[Table\left[Mod\left[11^{Mod[2^k,8530]},8777\right],\{k,2,62\}\right],2]}$$

$$0,1,1,0,1,0,1,0,0,0,0,1,1,1,0,0,1,0,1,1,1,0,1,1,0,1,0,1,0,0,1,0,$$

$$0,0,1,0,0,1,0,0,1,0,0,1,1,1,1,1,1,1,1,0,1,0,0,1,1,1,0,1,1,0.$$

Глава 2. Тестирование генераторов псевдослучайных чисел

2.1. Критерий согласия «хи-квадрат» К. Пирсона[11]

Пусть

$$B_1, B_2, \ldots, B_n \tag{2.1}$$

– последовательность результатов наблюдений n независимых повторений некоторого опыта, который заканчивается одним из k возможных исходов

$$A_1, \ldots, A_k.$$

Считая исходы случайными событиями, обозначим вероятности этих исходов через

$$p_i := \mathbb{P}\{A_i\},\ i = \overline{1,k}.$$

Будем считать, что $p_i > 0$ для любого $i = \overline{1,k}$, и $p_1 + \ldots + p_k = 1$.

Обозначим через $m_1(n), \ldots, m_k(n)$ число опытов, которые заканчиваются после n испытаний (повторений опыта), соответственно, исходами $A_1, \ldots, A_k$. Тогда

$$m_1(n) + \ldots + m_k(n) = n.$$

По закону больших чисел (в форме Бернулли) при $n \to \infty$ имеет место сходимость по вероятности *относительных частот*

$$\frac{m_1(n)}{n}, \ldots, \frac{m_k(n)}{n}$$

к *теоретическим вероятностям* $p_1, \ldots, p_k$:

$$\frac{m_i(n)}{n} \xrightarrow{\mathbb{P}} p_i, \quad \text{для любого } i = \overline{1,k}. \tag{2.2}$$

В самом деле, введем случайные величины

[11] Carl Pearson (1903 г.).

$X_{j\,i} := \begin{cases} 1, & \text{если в j-м испытании произошло случайное событие } A_i; \\ 0, & \text{если в j-м испытании не произошло случайное событие } A_i, \end{cases}$

тогда для любого $i = \overline{1,k}$.

$$\mathbb{P}\{X_{j\,i} = 1\} = \mathbb{P}\{A_i\} = p_i, \quad \mathbb{P}\{X_{j\,i} = 0\} = \mathbb{P}\{\overline{A_i}\} = 1 - p_i,$$

$$\frac{m_i(n)}{n} = \frac{1}{n}\sum_{j=1}^{n} X_{j\,i}.$$

Кроме того,

$$X_{1\,i}, X_{2\,i}, \ldots, X_{n\,i}$$

последовательность независимых[12] одинаково распределенных случайных величин, образующих схему Бернулли. Поэтому по закону больших чисел в форме Бернулли, при $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n} X_{j\,i} \xrightarrow{\mathbb{P}} p_i, \quad \text{для любого } i = \overline{1,k}.$$

Замечание. Можно утверждать даже большее: по усиленному закону больших чисел в форме Бореля при $n \to \infty$ имеет место сходимость почти наверное

$$\frac{1}{n}\sum_{j=1}^{n} X_{j\,i} \xrightarrow{\text{п.н.}} p_i, \quad \text{для любого } i = \overline{1,k}. \tag{2.3}$$

Для более углубленного исследования сходимостей (2.2) и (2.3) в качестве расстояния между k-мерными векторами относительных частот

$$\boldsymbol{m}_n = \left(\frac{m_1(n)}{n}, \ldots, \frac{m_k(n)}{n}\right)$$

и теоретических вероятностей

$$\boldsymbol{p} = (p_1, \ldots, p_k),$$

[12] Результаты испытаний независимы.

будем рассматривать случайную величину

$$\varrho\left(\boldsymbol{m}_n;\boldsymbol{p}\right) := \sqrt{\sum_{i=1}^{k}\left(\frac{m_i(n)}{n}-p_i\right)^2\frac{1}{p_i}}.$$

От евклидовой метрики

$$\sqrt{\sum_{i=1}^{k}\left(\frac{m_i(n)}{n}-p_i\right)^2}$$

метрика $\varrho\left(\boldsymbol{m}_n;\boldsymbol{p}_n\right)$ отличается лишь тем, что разные координаты векторов входят в эту метрику с различными весами $\frac{1}{p_i}$.

Из соотношений (2.3) следует, что при $n\to\infty$

$$\varrho\left(\boldsymbol{m}_n;\boldsymbol{p}\right) \xrightarrow{\text{п.н.}} 0. \tag{2.4}$$

Однако, если в формуле для $\varrho\left(\boldsymbol{m}_n;\boldsymbol{p}\right)$ вместо теоретических вероятностей

$$\boldsymbol{p} = (p_1,\ldots,p_k)$$

взять какой-либо другой набор ненулевых вероятностей

$$\boldsymbol{p}^\circ = (p_1^\circ,\ldots,p_k^\circ),\quad p_1^\circ+\ldots+p_k^\circ = 1,$$

то при $n\to\infty$

$$\varrho\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) \xrightarrow{\text{п.н.}} \sqrt{\sum_{i=1}^{k}\left(p_i-p_i^\circ\right)^2\frac{1}{p_i^\circ}} \;>\; 0. \tag{2.5}$$

Введем случайную величину

$$\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}\right) := n\sum_{i=1}^{k}\left(\frac{m_i(n)}{n}-p_i\right)^2\frac{1}{p_i}.$$

Так как

$$\varrho^2\left(\boldsymbol{m}_n;\boldsymbol{p}\right) = \frac{\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}\right)}{n},$$

то, ввиду соотношения (2.5), при $n \to \infty$

$$\chi_k^2\left(\boldsymbol{m}_n; \boldsymbol{p}^\circ\right) \xrightarrow{\text{п.н.}} +\infty.$$

Поскольку относительные частоты

$$\frac{m_1(n)}{n}, \ldots, \frac{m_k(n)}{n}$$

являются случайными величинами, то и введенные величины $\varrho\left(\boldsymbol{m}_n; \boldsymbol{p}\right)$ и $\chi_k^2\left(\boldsymbol{m}_n; \boldsymbol{p}\right)$ также являются случайными. Для описания поведения этих величин неплохо было бы знать их функции распределения.

В следующей теореме будет установлено, что случайная величина $\chi_k^2\left(\boldsymbol{m}_n; \boldsymbol{p}\right)$ (асимптотически при $n \to \infty$) имеет распределение «хи-квадрат» с $(k-1)$ степенями свободы.

Теорема (К. Пирсон). *Для любого* $x \geq 0$

$$\lim_{n\to\infty} \mathbb{P}\left\{\chi_k^2\left(\boldsymbol{m}_n; \boldsymbol{p}_n\right) \leq x\right\} = F_{k-1}(x),$$

где

$$F_{k-1}(x) = \frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})} \int\limits_0^x e^{-\frac{u}{2}}\, u^{\frac{k-1}{2}-1} du$$

распределение «хи-квадрат» с $(k-1)$ *степенями свободы.*

Доказательство с пояснениями можно найти в книге [4], стр. 321 – 327.

Справка. Напомним определение распределения «хи-квадрат»: *если случайная величина* Y *допускает представление вида*

$$Y = \sum_{i=1}^{m} X_i^2,$$

где X_i *– независимые, (0,1)-гауссовские случайные величины, то говорят, что* Y *имеет распределение «хи-квадрат» с* m *степенями свободы.*

Можно показать, что плотность распределения «хи-квадрат» с m степенями свободы имеет вид[13]

$$p_m(x) = \begin{cases} \frac{1}{2^{m/2}\Gamma(m/2)}\, e^{-x/2}\, x^{m/2-1}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Распределение «хи-квадрат» представляет собой частный случай гамма-распределения.

Проверка простой гипотезы

Будем считать, что теоретические вероятности

$$\boldsymbol{p} = (p_1, \ldots, p_k)$$

неизвестны. В этой ситуации можно использовать теорему К. Пирсона для проверки простой гипотезы о равенстве теоретических вероятностей некоторым гипотетическим (ненулевым) вероятностям

$$\boldsymbol{p}^\circ = (p_1^\circ, \ldots, p_k^\circ), \quad p_1^\circ + \ldots + p_k^\circ = 1.$$

Итак, проверяется простая гипотеза[14] H_0 :
выборка (2.1) *относится к известному распределению*

$$\boldsymbol{p}^\circ = (p_1^\circ, \ldots, p_k^\circ), \text{ т. е. : } \boldsymbol{p} = \boldsymbol{p}^\circ,$$

против альтернативной гипотезы H_1 :
выборка (2.1) *не относится к известному распределению*

$$\boldsymbol{p}^\circ = (p_1^\circ, \ldots, p_k^\circ), \text{ т. е. } \boldsymbol{p} \neq \boldsymbol{p}^\circ.$$

Для проверки гипотезы H_0 рассматривают *статистику «хи-квадрат» Пирсона для простой гипотезы*

$$\chi_k^2\,(\boldsymbol{m}_n; \boldsymbol{p}^\circ) = n \sum_{i=1}^{k} \left(\frac{m_i(n)}{n} - p_i^\circ \right)^2 \frac{1}{p_i^\circ}\,.$$

---

[13] См., например, Гнеденко Б.В. Курс теории вероятностей. М.: Наука, 1988.

[14] Буква H – первая буква английского слова *Hypothesis* (гипотеза).

Если гипотеза H_0 справедлива, то по теореме К. Пирсона

$$\lim_{n\to\infty} \mathbb{P}\left\{\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) \le x\right\} \to F_{k-1}(x).$$

Таким образом, в этом случае даже при достаточно больших n статистика $\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right)$ ограничена с положительной вероятностью.

Если же гипотеза H_0 несправедлива, то ввиду соотношения (2.5)

$$\lim_{n\to\infty} \chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) \overset{\text{п.н.}}{=} +\infty.$$

Гипотезу H_0 отвергают, если рассчитанное на основе выборки (2.1) значение статистики $\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right)$ является *нетипично* большим.

Именно задавая *уровень значимости*[15] α, найдем *критическое значения* $\lambda_n(\alpha)$, определяемое как решение уравнения

$$\mathbb{P}\left\{\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) \le \lambda_n(\alpha)\right\} = 1-\alpha.$$

Оценка скорости сходимости в теореме К. Пирсона (см. [8] с. 275, [9] с. 111) позволяет при выполнении следующих условий

$$n \ge 50 \text{ и } np_i \ge 5 \text{ для всех } i = \overline{1,k} \tag{2.6}$$

вполне удовлетворительно заменить $\lambda_n(\alpha)$ на близкую ей величину $\chi^2_{1-\alpha,\,k-1}$, которая определяется как решение уравнения

$$F_{k-1}(\chi^2_{1-\alpha,\,k-1}) = 1-\alpha.$$

Таким образом, при выполнении условий (2.6)

$$\mathbb{P}\left\{\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) \le \chi^2_{1-\alpha,\,k-1}\right\} \approx F_{k-1}(\chi^2_{1-\alpha,\,k-1}) = 1-\alpha$$

и

$$\mathbb{P}\left\{\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) > \chi^2_{1-\alpha,\,k-1}\right\} \approx \alpha. \tag{2.7}$$

[15] Например α =0,05, или 0,01, или 0,005. Уровни значимости обычно измеряют в процентах. Так, при $\alpha = 0,05$ говорят о 5-процентном уровне значимости.

При выборе малых значений уровня значимости α вероятность в соотношении (2.7) будет мала, поэтому событие

$$\left\{\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) > \chi^2_{1-\alpha,k-1}\right\}$$

практически невозможно. Таким образом, гипотезу $H_0 : \boldsymbol{p} = \boldsymbol{p}^\circ$, приводящую к появлению практически невозможного события, следует отвергнуть как практически неверную[16].

Алгоритм проверки гипотезы. Если рассчитанное по выборке (2.1) значение

$$\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) > \chi^2_{1-\alpha,\,k-1},$$

то (по критерию «хи-квадрат» с уровнем значимости α) гипотеза

$H_0 : \boldsymbol{p} = \boldsymbol{p}^\circ$ должна быть отвергнута.

В противном случае:

$$\chi_k^2\left(\boldsymbol{m}_n;\boldsymbol{p}^\circ\right) \le \chi^2_{1-\alpha,\,k-1},$$

гипотеза $H_0 : \boldsymbol{p} = \boldsymbol{p}^\circ$ принимается на уровне значимости α, по крайней мере до тех пор, пока не будут получены новые статистические данные.

Для $m = 4, 5, 6, 7$ приведем таблицу значений $\chi^2_{1-\alpha,\,m}$, соответствующих различным уровням значимости (см. [1] с. 167).

Число степеней свободы	*Уровни значимости*			
	0,1%	0,5%	1%	5%
4	18,467	14,860	13,277	9,488
5	20,515	16,750	15,086	11,070
6	22,458	18,548	16,812	12,592
7	24,322	20,278	18,475	14,067

Следует отметить, что принятие одной из двух гипотез на основе статистического критерия не означает, что принятая гипотеза истинна. Это говорит лишь о том, что принятая гипотеза лучше согласуется с данными наблюдений, чем отвергнутая.

[16] Это рассуждение представляет собой вероятностный вариант «метода доказательства от противного».

Замечание. В настоящее время для проверки согласия статистических данных предположениям о распределении случайных величин разработано довольно много различных критериев согласия. Поэтому выбор подходящего критерия и его использование для решения конкретной практической задачи требует не только хорошего знания соответствующих разделов математической статистики (смотри, например, [4], [5]), но и определенного практического опыта.

2.2. Тестирование равномерности

2.2.1. Обобщенный покер-тест

В обобщенном покер-тесте n-мерной равномерности рассматриваются серии равной длины[17] n :

$$\boldsymbol{x} = (x_1, \ldots, x_n).$$

Элементы x_i этих серий являются буквами алфавита длины

$$\mathcal{A} = \{\alpha_1, \ldots, \alpha_N\}.$$

Введем функцию $t = t(\boldsymbol{x})$, которая определяет количество различных букв, содержащихся в серии $\boldsymbol{x}$. Заметим, что максимальное число различных букв из алфавита $\mathcal{A}$, содержащихся в серии длины n, равно

$$L = \min(n, N).$$

Итак,

$$t(\boldsymbol{x}) = \begin{cases} 1; & \text{если } \boldsymbol{x} \text{ состоит из одинаковых элементов;} \\ 2; & \text{если } \boldsymbol{x} \text{ содержит ровно } 2 \text{ различных элемента;} \\ 3; & \text{если } \boldsymbol{x} \text{ содержит ровно } 3 \text{ различных элемента;} \\ \ldots. & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ L; & \text{если } \boldsymbol{x} \text{ содержит ровно } L \text{ различных элементов.} \end{cases}$$

[17] В классическом покер-тесте рассматриваются серии длины $n = 5$.

Теорема. Число различных серий $\boldsymbol{x}$, для которых

$$t(\boldsymbol{x}) = m,\ (m = \overline{1, L}\,),$$

равно

$$N(N-1)\dots(N-m+1)S(n,m).$$

Вероятность события, состоящего в том, что для «наугад» выбранной серии $\boldsymbol{x}$ функция $t(\boldsymbol{x}) = m,\ (m = \overline{1, L}\,)$ вычисляется по формуле

$$\mathbb{P}\{\boldsymbol{x} : t(\boldsymbol{x}) = m\} = \frac{|K_m|}{|\Omega|} = \frac{(N-1)\dots(N-m+1)S(n,m)}{N^{n-1}}, \quad (2.8)$$

где $S(n,m)$ – число Стирлинга второго рода.

Доказательство. Вначале рассмотрим множество всех серий

$$\Omega = \{\boldsymbol{x} = (x_1, \dots, x_n) : x_i \in \mathcal{A},\ i = \overline{1, n}\,\}.$$

Нетрудно видеть, что

$$|\Omega| = N^n.$$

Рассмотрим подмножества $K_m \subset \Omega,\ (m = \overline{1, L})$, состоящие из серий $\boldsymbol{x} = (x_1, \dots, x_n)$, которые содержат ровно m различных букв:

$$K_m = \{\boldsymbol{x} \in \Omega :\ t(\boldsymbol{x}) = m\}.$$

Тогда

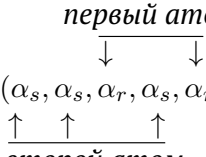
$$\Omega = \bigcup_{m=1}^{L} K_m, \text{ и } K_i \cap K_j = \varnothing, \text{ при } i \neq j.$$

Дадим описание атомов разбиения множества K_m. Множество номеров $\{1, \dots, n\}$ элементов каждой серии

$$(x_1, \dots, x_n) \in K_m$$

разобьем на m непустых атомов следующим образом. Вначале в каждый атом помещаем номер одного из m различных элементов, а затем оставшиеся места в атомах заполняются номерами

элементов, совпадающих с первоначально выбранными. Другими словами, элементы x_i с номерами i, принадлежащими одному атому, являются одинаковыми буквами алфавита $\mathcal{A}$. Если же номера i, j элементов x_i и x_j принадлежат разным атомам, то эти элементы являются разными буквами алфавита $\mathcal{A}$.



$$\{1,2,3,4,5\} = \{1,2,4\} \cup \{3,5\}$$

Пример разбиения 5-элементной серии на два атома.

Число таких (неупорядоченных) разбиений равно числу Стирлинга второго рода $S(n,m)$. Теперь осталось заметить, что букву $\alpha_{l_1} \in \mathcal{A}$, с которой совпадают элементы с индексами из первого атома, можно выбрать N способами, вторую букву $\alpha_{l_2} \in \mathcal{A}$, с которой совпадают элементы с индексами из второго атома, можно выбрать $N-1$ способом и т.д., наконец, букву $\alpha_{l_m} \in \mathcal{A}$, с которой совпадают элементы с индексами из m-го атома, можно выбрать $N-m+1$ способом. Следовательно, общее число серий, $(x_1, \ldots, x_n)$, которые соответствуют одному разбиению множества номеров $\{1, \ldots, n\}$, равно

$$N(N-1)\ldots(N-m+1).$$

Так как число таких разбиений равно $S(n,m)$, то число всех серий

$$(x_1, \ldots, x_n) \in K_m$$

равно

$$N(N-1)\ldots(N-m+1)S(n,m).$$

Используя формулу классической вероятности, можно утверждать, что

$$\mathbb{P}\{\boldsymbol{x} : t(\boldsymbol{x}) = m\} = \frac{|K_m|}{|\Omega|} = \frac{(N-1)\ldots(N-m+1)S(n,m)}{N^{n-1}}.\blacksquare$$

Найденные в предыдущей теореме вероятности (2.8)

$$p_m := \mathbb{P}\{\boldsymbol{x} : t(\boldsymbol{x}) = m\}, \quad m = \overline{1, L},$$

называются *теоретическими.* Так как вычисление этих вероятностей основано на формуле классической вероятности, то серии

$$\boldsymbol{x} = (x_1, \ldots, x_n)$$

являются реализациями *независимых, равномерно распределенных случайных величин*

$$\mathbf{X} = (X_1, \ldots, X_n),$$

которые принимают свои значения в алфавите $\mathcal{A}$.

Эмпирические вероятности (частоты) находятся с помощью псевдослучайной последовательности вырабатываемой ГПСЧ

$$x_1,\, x_2,\, \ldots,\, x_{r+k-1},$$

из которой (с помощью сдвига на r индексов) выделяется k серий

$$\underbrace{x_1,\, \ldots,\, x_r}_{1\,\text{серия}};\ \underbrace{x_2,\, \ldots,\, x_{r+1}}_{2\,\text{серия}};\ \ldots;\ \underbrace{x_k,\, \ldots,\, x_{r+k-1}}_{k\,\text{серия}}.$$

Частота серий, в которых $t(\boldsymbol{x}) = m$, определяются по формуле

$$p^{\circ}_m = \frac{\text{число серий, в которых } t(\boldsymbol{x}) = m}{k}, \quad m = \overline{1, L}.$$

Наконец, сравнение теоретических вероятностей с эмпирическими (частотами)

$$\boldsymbol{p} = (p_1,\, p_2,\, \ldots,\, p_n), \quad \boldsymbol{p}^{\circ} = (p^{\circ}_1,\, p^{\circ}_2,\, \ldots,\, p^{\circ}_n)$$

осуществляется с помощью критерия К. Пирсона.

2.2.2. Тест «собирателя купонов»

В тесте «собирателя купонов» n-мерной равномерности рассматриваются серии длины r :

$$\boldsymbol{x}^{(r)} = (x_1, \ldots, x_r).$$

Элементы x_i этих серий являются буквами алфавита длины N $(r \geq N)$:

$$\mathcal{A} = \{\alpha_1, \ldots, \alpha_N\},\ (\textit{купоны}).$$

Рассмотрим множество всех серий длины r

$$\Omega = \{\boldsymbol{x}^{(r)} = (x_1, \ldots, x_r) : x_i \in \mathcal{A},\ i = \overline{1, r}\,\}.$$

Нетрудно видеть, что

$$|\,\Omega| = N^{\,r}.$$

Найдем вероятность события, состоящего в том, что серия (минимальной) длины r содержит все буквы алфавита $\mathcal{A}$. (На r-м шаге происходит *первый* сбор всех «купонов».) Множество всех таких серий обозначим через M_r. Итак, нам нужно вычислить вероятность

$$\mathbb{P}\{\boldsymbol{x}^{(r)} \in M_r\}.$$

Теорема. Вероятность того, что на r-м шаге происходит *первый* сбор всех «купонов», вычисляется по формуле

$$p_r := \mathbb{P}\{\boldsymbol{x}^{(r)} \in M_r\} = \frac{(N-1)!}{N^{\,r-1}}\, S(r-1, N-1),\ r = \overline{N, +\infty}. \quad (2.9)$$

Доказательство. Вначале найдем вероятность того, что серия длины r содержит все буквы алфавита $\mathcal{A}$. Для этого подсчитаем число серий

$$\boldsymbol{x}^{(r)} = (x_1, \ldots, x_r),$$

которые содержат все буквы алфавита $\mathcal{A}$. Это можно сделать следующим образом. Множество номеров $\{1, \ldots, r\}$ элементов каждой серии

$$(x_1, \ldots, x_r) \in \Omega$$

разобьем на N непустых атомов следующим образом. В первый атом поместим индексы (номера) первой (произвольно выбранной) буквы алфавита. Это можно сделать N способами. Во второй атом поместим индексы (номера) второй (произвольно выбранной) буквы алфавита. Это можно сделать $N-1$ способами. И так далее, в последний N-й атом поместим индексы (номера)

последней N-й буквы алфавита. Это можно сделать только одним способом. Таким образом, одно разбиение множества номеров $\{1,\dots,r\}$ на N атомов порождает $N!$ различных серий

$$\{\boldsymbol{x}^{(r)} = (x_1,\dots,x_r) : x_i \in \mathcal{A},\ i = \overline{1,r}\,\},$$

которые содержат все буквы алфавита $\mathcal{A}$. Тогда совокупность всех разбиений множества номеров $\{1,\dots,r\}$ на N атомов порождает

$$N!\,S(r,N)$$

различных серий, которые содержат все буквы алфавита $\mathcal{A}$. Следовательно, вероятность события, состоящего в том, что серия длины r содержит все буквы алфавита $\mathcal{A}$, равна

$$\frac{N!\,S(r,N)}{N^{\,r}}.$$

Отсюда вероятность дополнительного события, состоящего в том, что серия длины r не содержит все буквы алфавита $\mathcal{A}$, равна

$$q_r := 1 - \frac{N!\,S(r,N)}{N^{\,r}}. \tag{2.10}$$

Обозначим через

$$L_k := \{\boldsymbol{x}^{(k)} = (x_1,\dots,x_k) : x_i \in \mathcal{A},\ i = \overline{1,k}\,\}$$

множество всех серий длины k, которые *не содержат всех букв алфавита* $\mathcal{A}$. Выберем произвольную серию

$$\boldsymbol{x}^{(r-1)} = (x_1,\dots,x_{r-1}) \in L_{r-1}$$

и добавим к ней произвольный элемент $\widetilde{x}_r \in \mathcal{A}$. Введем обозначения для такой «раздутой» серии

$$(\boldsymbol{x}^{(r-1)},\widetilde{x}_r) = (x_1,\dots,x_{r-1},\widetilde{x}_r)$$

и для множества всех таких серий

$$\widetilde{L}_r = \{(\boldsymbol{x}^{(r-1)},\widetilde{x}_r) : \boldsymbol{x}^{(r-1)} \in L_{r-1}, \widetilde{x}_r \in \mathcal{A}\}.$$

Нетрудно видеть, что

$$|\widetilde{L}_r| = N \cdot |L_{r-1}|.$$

Далее, для каждой «раздутой» серии должны выполняться две исключающие друг друга возможности:

$$(\boldsymbol{x}^{(r-1)}, \widetilde{x}_r) \in L_r, \text{ или } (\boldsymbol{x}^{(r-1)}, \widetilde{x}_r) \in M_r.$$

Поэтому

$$\widetilde{L}_r = L_r \cup M_r, \text{ причем } L_r \cap M_r = \varnothing.$$

Следовательно,

$$M_r = \widetilde{L}_r \setminus L_r,$$

и

$$|M_r| = |\widetilde{L}_r| - |L_r| = N \cdot |L_{r-1}| - |L_r|.$$

Тогда

$$\mathbb{P}\{\boldsymbol{x}^{(r)} \in M_r\} = \frac{N \cdot |L_{r-1}| - |L_r|}{N^r} = \frac{|L_{r-1}|}{N^{r-1}} - \frac{|L_r|}{N^r}. \tag{2.11}$$

Используя ранее введенные обозначения (2.10), можно утверждать, что

$$q_r = \frac{|L_r|}{N^r}.$$

Поэтому, ввиду формул (2.10) и (2.11), считая $r \geq N+1$,

$$\mathbb{P}\{\boldsymbol{x}^{(r)} \in M_r\} = q_{r-1} - q_r = \frac{N!\,S(r,N)}{N^{\,r}} - \frac{N!\,S(r-1,N)}{N^{\,r-1}} =$$

$$= \frac{N!}{N^{\,r}}\left[S(r,N) - N\,S(r-1,N)\right].$$

Отсюда, используя рекуррентное соотношение для чисел Стирлинга второго рода[18] , будем иметь

$$\mathbb{P}\{\boldsymbol{x}^{(r)} \in M_r\} = \frac{(N-1)!}{N^{\,r-1}}\,S(r-1,N-1),\ r = \overline{N+1,+\infty}. \tag{2.12}$$

[18]См. Приложение.

В случае, когда длина серии $r = N$, множество серий L_{N-1} состоит из всех серий длины $N-1$, поэтому $|L_{N-1}| = N^{N-1}$. Таким образом, ввиду формул (2.10) и (2.11)

$$\mathbb{P}\{\boldsymbol{x}^{(N)} \in M_N\} = 1 - 1 + \frac{N!\,S(N,N)}{N^N} = \frac{N!}{N^N} = \frac{(N-1)!}{N^{N-1}}. \quad (2.13)$$

Последнее равенство сразу следует из классического определения вероятности.

Заметим, что формулы (2.12) и (2.13) объединяются в одну формулу

$$\mathbb{P}\{\boldsymbol{x}^{(r)} \in M_r\} = \frac{(N-1)!}{N^{\,r-1}}\,S(r-1,N-1),\ r = \overline{N,+\infty}. \blacksquare \quad (2.14)$$

С помощью формулы (2.9) подсчитываются вероятности

$$p_N,\, p_{N+1},\, \ldots, p_{N+t-1},$$

первого сбора всех купонов на N-м шаге, на $N+1$-м шаге, и т.д., на шаге с номером $N+t-1$. Кроме этого, используя формулу (2.10), подсчитывается вероятность того события, что серия длиной $N+t-1$ не содержит всех букв алфавита

$$q_{N+t-1} = 1 - \frac{N!S(N+t-1,N)}{N^{N+t-1}}.$$

Найденные вероятности

$$p_N,\, p_{N+1},\, \ldots, p_{N+t-1}, q_{N+t-1}$$

называются *теоретическими.* Так как вычисление этих вероятностей основано на формуле классической вероятности, то серии

$$\boldsymbol{x}^{(N+t-1)} = (x_1, \ldots, x_{N+t-1})$$

предполагаются реализациями *независимых, равномерно распределенных случайных величин*

$$\mathbf{X} = (X_1, \ldots, X_r),$$

которые принимают свои значения в алфавите $\mathcal{A}$.

Эмпирические вероятности (частоты событий, состоящих в том, что первый сбор всех купонов происходит на r-м шаге) находятся с помощью псевдослучайной последовательности вырабатываемой ГПСЧ

$$x_1, x_2, \ldots, x_{r+k-1}.$$

Из этой последовательности (с помощью сдвига на r индексов) выделяется k серий

$$\underbrace{x_1, \ldots, x_r}_{1\,\text{серия}}; \underbrace{x_2, \ldots, x_{r+1}}_{2\,\text{серия}}; \ldots; \underbrace{x_k, \ldots, x_{r+k-1}}_{k\,\text{серия}}.$$

Частота серий, в которых на r-м шаге происходит первый сбор всех купонов, определяется по формуле

$$p_r^\circ = \frac{\text{число серий, в которых на r-м шаге происходит первый сбор всех купонов}}{k}.$$

Эти вычисления проводятся для каждого $r = \overline{N, N+t-1}$.

Аналогичным образом, используя псевдослучайную последовательность, вырабатываемую ГПСЧ, подсчитывается q°_{N+t-1} - частота события: *серия длиной* $N+t-1$ *не содержит всех букв алфавита*.

Далее сравнение теоретических вероятностей с эмпирическими (частотами) осуществляется с помощью критерия К. Пирсона. $\boldsymbol{p} = (p_N, p_{N+1}, \ldots, p_{N+t-1}, q_{N+t-1})$, $\boldsymbol{p}^\circ = (p_N^\circ, p_{N+t-1}^\circ, \ldots, p_{N+t-1}^\circ, q_{N+t-1}^\circ)$

В настоящее время для оценки качества случайных и псевдослучайных последовательностей, помимо покер-теста и теста собирателей купонов, используются большое число других тестов. Известны несколько семейств статистических тестов, которые называют «батареями тестов». Например, «батарея тестов Д. Кнута (*D. Ktuth*)», «батарея тестов Дж. Марсальи (*G. Marsaglia*)», «батарея тестов Национального института стандартов США (*NIST*)» (см. [1], [2], [3]).

Существование большого числа тестов связано с тем обстоятельством, что с помощью каждого теста можно заметить лишь вполне определенные виды «отклонений» свойств случайных и псевдослучайных последовательностей от соответствую-

щих свойств идеальных случайных последовательностей. Универсального теста, который мог бы зафиксировать все виды указанных отклонений, в настоящее время не существует.

Кроме того, для проверки качества генераторов случайных и псевдослучайных чисел, наряду с критерием К. Пирсона, используются критерий А.Н. Колмогорова и другие критерии согласия (см. [2], [4], [5]).

Глава 3. Применение генераторов псевдослучайных чисел

3.1. Криптографическая защита информации

Генераторы случайных и псевдослучайных чисел широко используются в различных методах защиты информации. Начнем с использования этих генераторов в криптографиии.

Шифр простой замены

Шифрование по методу простой замены сводится к созданию по определённому алгоритму таблицы шифрования, в которой каждой букве открытого текста сопоставляется единственная буква шифртекста. Само шифрование заключается в замене букв согласно таблице. Для расшифровки достаточно иметь ту же таблицу либо знать алгоритм, по которому она генерируется.

Частотный метод вскрытия шифра простой замены основан на том обстоятельстве, что частотные характеристики букв, а также подряд идущих биграмм и триграмм шифртекста и открытого текста одинаковы.

Таблица частот букв русского языка

32-буквенный алфавит (добавлен пробел "-", Е=Ё, Ь=Ъ)

-	О	Е,Ё	А	И	Т	Н	С
0.175	0.090	0.072	0.062	0.062	0.053	0.053	0.045
Р	В	Л	К	М	Д	П	У
0.040	0.038	0.035	0.028	0.026	0.025	0.023	0.021
Я	Ы	З	Ь,Ъ	Б	Г	Ч	Й
0.018	0.016	0.016	0.014	0.014	0.013	0.012	0.010
Х	Ж	Ю	Ш	Ц	Щ	Э	Ф
0.009	0.007	0.006	0.006	0.004	0.003	0.003	0.002

Источник данных: Яглом А. М., Яглом И. М. *Вероятность и информация.* — М.: Наука, 1973.

1. *Подсчет числа встречаемости букв шифртекста и вычисление их частот.*

2. *Подсчет числа встречаемости биграмм и триграмм подряд идущих букв шифртекста и вычисление их частот.*

Если длина шифртекста достаточно велика, то найденные в пунктах 1 и 2 частоты шифртекста окажутся близкими к известным значениям этих частот для русского языка. При этом учитывается наличие предпочтительных связей каждой буквы русского языка с остальными буквами[19]. На этой основе происходит отождествление букв шифротекста с буквами русского языка. Т.е. осуществляется дешифрование (вскрытие) шифртекста.

Абсолютно случайные последовательности

В этом параграфе будут рассматриваться случайные величины, принимающие значения во множестве $\mathbb{Z}_\nu = \{0, 1, \ldots, \nu - 1\}$.

Определение. *Абсолютно случайными последовательностями будем называть последовательности дискретных случайных величин* $\{X_t\}$ $(t = \overline{1, \infty})$*, которые принимают значения во множестве* $\mathbb{Z}_\nu$ *и обладают следующими свойствами:*

1) *для любого натурального* n *и для любых индексов* $1 \leq t_1 < \ldots < t_n < \infty$ *случайные величины* $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ *независимы в совокупности;*

2) $\mathbb{P}\{X_t = i\} \equiv \frac{1}{\nu}$, *для любого* $i = \overline{0, \nu - 1}$ *и для любого* $t = \overline{1, \infty}$.

Абсолютно случайные последовательности называют также *равномерно распределенными случайными последовательностями (РРСП)* или *чисто случайными последовательностями*.

Теорема. *Если* $\{X_t\}$ *– абсолютно случайная последовательность, то для любого натурального* n *и для любых индексов* $1 \leq t_1 < \ldots < t_n < \infty$ *распределение вероятностей случайного вектора*

$$(X_{t_1}, \ldots, X_{t_n})$$

[19] См. [6], стр. 434 – 447.

является равномерным:

$$\mathbb{P}\{X_{t_1} = i_1, \ldots, X_{t_n} = i_n\} = \frac{1}{\nu^n},\ i_1, \ldots, i_n \in \mathbb{Z}_\nu.$$

Доказательство. По свойствам 1 и 2

$$\mathbb{P}\{X_{t_1} = i_1, \ldots, X_{t_n} = i_n\} = \mathbb{P}\{X_{t_1} = i_1\} \cdot \ldots \cdot \mathbb{P}\{X_{t_n} = i_n\} = \frac{1}{\nu^n}. \blacksquare$$

Через $\oplus_\nu$ и $\ominus_\nu$ будем обозначать операции сложения и вычитания по модулю ν:

$$a \oplus_\nu b = c \Leftrightarrow a = c \ominus_\nu b, \text{ для } a, b, c \in \mathbb{Z}_\nu.$$

Теорема. *Если $\{X_m\}$ – абсолютно случайная последовательность, то для любой неслучайной последовательности $\{u_m\}$, элементы которой принимают значения из множества $\mathbb{Z}_\nu$, последовательность $\{X_m \oplus_\nu u_m\}$ является абсолютно случайной.*

Доказательство. Ясно, что свойство взаимной независимости при покоординатном сложении по модулю ν сохраняется. Действительно, для любой неслучайной последовательности $(d_1, d_2, \ldots, d_m, \ldots)$, элементы которой принимают значения из множества $\{0, 1, \ldots, \nu - 1\}$, имеют место равенства

$$\mathbb{P}\{X_1 \oplus_\nu u_1 = d_1; \ldots; X_m \oplus_\nu u_m = d_m\} =$$

$$= \mathbb{P}\{X_1 = d_1 \ominus_\nu u_1; \ldots; X_m = d_m \ominus_\nu u_m\} =$$

$$= \mathbb{P}\{X_1 = d_1 \ominus_\nu u_1\} \cdot \ldots \cdot \mathbb{P}\{X_m = d_m \ominus_\nu u_m\} =$$

$$= \mathbb{P}\{X_1 \oplus_\nu u_1 = d_1\} \cdot \ldots \cdot \mathbb{P}\{X_m \oplus_\nu u_m = d_m\}.$$

Кроме того, для любого $m = \overline{1, \infty}$

$$\mathbb{P}\{X_m \oplus_\nu u_m = d_m\} = \mathbb{P}\{X_m = d_m \ominus_\nu u_i\} = \frac{1}{\nu}. \blacksquare$$

Таким образом, поэлементное сложение по модулю ν членов последовательности

$$d_1, \ldots, d_m, \ldots \quad (\textit{открытый текст})$$

с соответствующими членами абсолютно случайной последовательности

$$X_1, \ldots, X_m, \ldots \quad (\textit{ключ})$$

снова приводит к абсолютно случайной последовательности

$$d_1 \oplus_\nu X_1, \ldots, d_m \oplus_\nu X_m, \ldots, \quad (\textit{шифртекст})$$

у которой равные вероятности появления различных букв алфавита и нет зависимости между различными элементами. Частотный метод вскрытия для таких последовательностей становится невозможным.

Поточные шифры

При практической реализации теоремы, рассмотренной в предыдущем разделе, вместо абсолютно случайных последовательностей *случайных величин* используют последовательности случайных или псевдослучайных *чисел*.

Шифр, называемый *одноразовым блокнотом*, использует секретный ключ (с длиной, совпадающей с длиной открытого текста), который состоит из букв открытого алфавита, выбираемых совершенно случайно, причем каждый ключ используется только один раз.

Хотя одноразовый блокнот обладает свойством совершенной криптостойкости,[20] требования к ключу одноразового блокнота приводят к определенным неудобствам: слишком длинный ключ, который разрешается использовать для зашифрования только одного открытого текста.

Поточный шифр фактически имитирует одноразовый блокнот, используя короткий ключ для генерации достаточно длинной псевдослучайной шифрующей последовательности (см. [7]). Например, часто применяется линейный регистр сдвига с обратной связью (LFSR) . В этом случае роль ключа играет секретный короткий список начальных состояний LFSR, которые выбираются случайно. В качестве шифрующей последовательности используется последовательность псевдослучайных чисел, генерируемая LFSR.

[20] См. [2], стр. 146.

На комбинации нескольких регистров сдвига с линейно обратной связью основаны такие поточные шифры, как A5/1 и A5/2, используемые в стандарте GSM, и шифр E0, используемый в Bluetooth.

Изучение того, каким образом генерировать псевдослучайные последовательности, составляет основную часть разработок в области поточных шифров. Шифрующая последовательность порождается генератором псевдослучайных чисел (ГПСЧ) или, как выражаются криптографы, *генератором гаммы*. Достоинством поточных шифров является высокое быстродействие.

3.2. Генераторы шума

Вначале дадим краткое описание использования генераторов случайных чисел в системах виброакустической защиты помещений[21].

Акустические волны, создаваемые человеческой речью, воздействуют на различные конструкции помещения (перегородки, стены, перекрытия, окна, двери) и инженерные системы (трубопроводы, вентиляционные каналы и прочее), передавая им часть своей энергии.

Возникающие в результате этих воздействий слабые колебания конструкций и инженерных систем могут быть приняты и усилены специальными приборами (например, электронными стетоскопами или лазерными микрофонами).

Генератор шума создаёт специальную шумовую помеху (например, так называемый «белый» шум), которая передается строительным и инженерным конструкциям и системам. Наличие такой шумовой помехи в акустическом диапазоне позволяет устранить возможные утечки информации по техническим каналам и обеспечить акустическую «непроницаемость» защищаемых помещений.

Важной характеристикой генератора шума является коэффициент качества шума, показывающий меру сходства генери-

[21] bib.pps.ru/wiki/Хорошо шумим.

руемой помехи с «белым» шумом. Чем выше значение коэффициента качества шума, тем труднее выделить скрываемый речевой сигнал при использовании различных методов шумоочистки.

Аналогичную функцию выполняют генераторы шума (помех), которые предназначены для маскировки информативных побочных электромагнитных излучений и наводок (ПЭМИН) персональных компьютеров, рабочих станций компьютерных сетей и комплексов на объектах вычислительной техники[22].

Такие генераторы излучают в окружающее пространство электромагнитное поля шума (ЭМПШ) в определенном диапазоне частот.

Главная задача генератора шума — это полная защита информации по каналам ПЭМИН и по слабо защищенным беспроводным сетям.

Такие генераторы можно использовать: на экзаменах в школах и вузах, в театрах и других общественных местах (где важен момент отключения телефонов), для защиты от прослушивания на важных встречах и переговорах.

Большинство из подслушивающих «жучков» работают на частотах, которые доступны генераторам помех. Поэтому, установив такое устройство у себя в кабинете, можно спокойно говорить о своих делах, не опасаясь разглашения конфиденциальной информации.

Мобильные генераторы помех можно использовать в автомобилях, если у вас есть обоснованные подозрения, что вас отслеживают с помощью ГЛОНАСС-трекеров, GPS; с помощью специальных программ, установленных в вашем мобильном телефоне.

3.3. Метод Монте-Карло

Кроме задач защиты информации, ГПСЧ широко используются в методах Монте-Карло.

[22]www.suritel.ru

Приведем простейший вариант метода Монте-Карло для приближенного вычисления определенных интегралов

$$\int_a^b f(x)dx$$

от функций $f(x)$, которые нельзя (или очень трудно) интегрировать аналитически[23].

Пусть X – случайная величина, имеющая равномерное распределение на отрезке $[a,b]$ с плотностью

$$u(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b], \\ 0, & x \notin [a,b]. \end{cases}$$

Тогда математическое ожидание случайной величины $(b-a)f(X)$ вычисляется по формуле

$$\mathbb{M}\{(b-a)f(X)\} = (b-a)\int\limits_{-\infty}^{\infty} f(x)u(x)dx =$$

$$= (b-a)\cdot\frac{1}{(b-a)}\int\limits_a^b f(x)dx = \int\limits_a^b f(x)dx.$$

Рассматривая независимые и равномерно распределенные на отрезке $[a,b]$ случайные величины

$$X_1,\, X_2,\, \ldots X_n, \tag{3.1}$$

по (усиленному) закону больших чисел будем иметь сходимость с вероятностью единица выборочного среднего к математическому ожиданию

$$\mathbb{P}\left\{\lim_{n\to\infty}\frac{b-a}{n}\sum_{k=1}^{n} f(X_k) \;=\; \int\limits_a^b f(x)dx\right\} = 1.$$

[23]Метод Монте-Карло обычно применяют для вычисления *кратных* интегралов. Кроме того, этот метод используют для решения линейных и нелинейных уравнений, задач поиска экстремальных значений функций многих переменных, а также других задач.

Заменяя идеальную последовательность случайных величин (3.1) псевдослучайной последовательностью

$$x_1,\, x_2,\, \ldots x_n,$$

выработанную генератором псевдослучайных чисел, для достаточно больших n получаем приближенное равенство

$$\frac{b-a}{n}\sum_{k=1}^{n} f(x_k) \approx \int_a^b f(x)dx.$$

Пример[24]. Рассмотрим интеграл

$$\int_0^{\pi} \sin(x)dx,$$

который, как известно, равен 2. Обозначим выборочное среднее

$$Y(n) := \frac{\pi}{n}\sum_{k=1}^{n} \sin(x_k).$$

Тогда

n	10	20	40	80	160
$Y(n)$	$2,213$	$1,191$	$1,948$	$1,989$	$1,993$

.

[24] См. [8], стр. 120.

Приложение

Числа Стирлинга второго рода

Договоримся обозначать число элементов (мощность) конечного множества W через $|W|$. Например, $|\{1,2,3,4\}| = 4$.

Определение. *Разбиением конечного множества W с $|W| = n$ называется система его подмножеств*

$$A_1, \ldots, A_k,\ 1 \le k \le n,$$

которые обладают свойствами

$$A_i \cap A_j = \varnothing,\ \textit{когда}\ i \ne j;\ \textit{и}\ \bigcup_{i=1}^{k} A_i = W. \qquad (n_1)$$

Подмножества A_i называются атомами разбиения, а числа $|A_i|$– мощностями атомов.

Различают два типа разбиений множеств: *неупорядоченные* и *упорядоченные*[25]. Два неупорядоченных разбиения конечного множества считаются равными, если они совпадают по составу, т.е. каждый атом одного разбиения является атомом другого и наоборот[26].

Рассмотрим задачу о нахождении общего числа $S(n,k)$ всех неупорядоченных разбиений n-элементного множества $W = \{a_1, \ldots, a_n\}$ на k произвольных *непустых* атомов

$$\{A_1, \ldots, A_k\},\ 1 \le k \le n.$$

Мощности атомов $|A_i| = r_i$ являются натуральными числами $(r_i \ge 1)$, для которых, ввиду свойств (n_1), выполняется соотношение

$$\sum_{i=1}^{k} r_i = n.$$

[25] В англоязычной литературе для упорядоченных разбиений используют термин *division*, а для неупорядоченных – *partition*.

[26] Два упорядоченных разбиения считаются равными, если они совпадают по составу и по порядку следования атомов. Заметим, что исходное множество W и, соответственно, все его подмножества (атомы) A_i мы считаем неупорядоченными.

Числа $S(n,k)$, где $1 \le k \le n$, называются *числами Стирлинга второго рода*.

Положим по определению

$$S(n,0) = 0 \text{ при } n \ge 1 \text{ и } S(0,k) = 0 \text{ при } k \ge 1.$$

Нетрудно видеть, что $S(n,k)$ равно числу способов размещения n *различных* предметов по k *одинаковым* ячейкам, при которых ни одна из них не останется пустой.

Примеры.

1. $W = \{a_1\},\ S(1,1) = 1.$

2. $W = \{a_1, a_2\},\ S(2,1) = 1,$ т.к. разбиение только одно: $\{a_1, a_2\}$;

$S(2,2) = 1,$ т.к. разбиение только одно: $\{\{a_1\}, \{a_2\}\}$.

3. $W = \{a_1, a_2, a_3\},\ S(3,1) = 1,$ т.к. разбиение только одно: $\{a_1, a_2, a_3\}$;

$S(3,2) = 3,$ т.к. в этом случае три разбиения:

$\{\{a_1\}\{a_2, a_3\}\}, \{\{a_2\}, \{a_1, a_3\}\}, \{\{a_3\}, \{a_1, a_2\}\}$;

$S(3,3) = 1,$ т.к. разбиение только одно: $\{\{a_1\}, \{a_2\}, \{a_3\}\}$.

Свойства чисел Стирлинга второго рода

1. *Числа Стирлинга второго рода удовлетворяют рекуррентному соотношению*

$$S(n,k) = S(n-1,k-1) + kS(n-1,k) \text{ при условии } S(0,0) = 1.$$

2. $S(n,2) = 2^{n-1} - 1.$

3. $S(n,k) = \frac{1}{k!} \sum\limits_{m=0}^{k} (-1)^{k-m} C_k^m\, m^n,$ где $C_k^m = \frac{m!}{k!(m-k)!}$.

Доказательства приведены в книге Стенли Р. *Перечислительная комбинаторика*. М.: Мир, 1990. 440 с.

Таблица чисел Стирлинга второго рода $S(n,k)$

$n \setminus k$	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	1	3	1							
4	1	7	6	1						
5	1	15	25	10	1					
6	1	31	90	65	15	1				
7	1	63	301	350	140	21	1			
8	1	127	966	1701	1050	266	28	1		
9	1	255	3025	7770	6951	2646	462	36	1	
10	1	511	9330	34105	42525	22827	5880	750	45	1

Исторические замечания

Карл Пирсон[27] (Carl Pearson 1857–1936 гг.) Окончил Кембриджский университет в 1879 году. Затем изучал физику в Гейдельбергском и Берлинском университетах. С 1884 по 1911 год – профессор прикладной математики и механики Лондонского университета, с 1911 года – директор Лаборатории евгеники Лондонского университета.

В 1896 году был избран членом Королевского общества, в 1898 году награждён Медалью Дарвина. В 1900 году основал журнал «Biometrika», посвящённый применению статистических методов в биологии.

Опубликовал основополагающие труды по математической статистике (более 400 работ), посвященные теории корреляции, критериям согласия, алгоритмам принятия решений и оценкам параметров.

Джеймс Стирлинг[28] (James Stirling 1692–1770 гг.) хорошо известен как автор асимптотической формулы

$$n! = \sqrt{2\pi n}\ n^n e^{-n}(1 + o(1)).$$

[27] http://ru.wikipedia.org (Пирсон, Карл.)

[28] Юшкевич А.П. История математики с древнейших времен до начала 19 столетия. Т. 3. М.: Наука, 1972.

Уроженец Шотландии, учился в Оксфордском университете. Из-за участия в политической деятельности, враждебной правящей в те годы в Англии династии (восстание якобитов-католиков), был вынужден эмигрировать в Венецию, где зарабатывал на жизнь частными уроками.

В 1725 г. с помощью И. Ньютона Дж. Стирлинг вернулся в Англию. В 1735 г. получил пост директора горного предприятия в Шотландии, который занимал до конца своей жизни.

Дж. Стирлинг имел научные контакты с А. де Муавром, который, будучи гугенотом, был вынужден эмигрировать из Франции в Англию после отмены Людовиком XIV Нантского эдикта в 1685 году. Гугенотами во Франции называли приверженцев кальвинизма. Одно из положений кальвинизма: успех в профессиональной деятельности человека служит подтверждением его избранности.

Само название «числа Стирлинга» стало использоваться в математике только с 1906 г.

Метод Монте-Карло был разработан в годы Второй мировой войны Дж. фон Нейманом, С. Уламом и Н. Метрополисом во время их работы в Манхэттенском проекте (создание атомной бомбы в США).

Станислав Улам пишет в своей автобиографии[29], что название метода было предложено Николасом Метрополисом в честь его дяди, который был азартным игроком.

В докомпьютерную эпоху именно рулетка (казино) являлась одним из немногих (наряду с бросанием монет или игральных костей, а также извлечением игральных карт) генераторов случайных чисел. В Монте-Карло (г. Монако) расположено одно из самых известных казино в Европе.

Шифры простой замены были известны с античных времен (см., например, *шифр Цезаря*, *квадрат Полибия*). Частотный анализ шифра простой замены был известен арабам, начиная с девятого века.

Одним из первых примеров поточных шифров являлся *шифр Вернама*, или «одноразовый блокнот» ("one-time pad").

[29]Улам С. Приключения математика. Ижевск: НИЦ РХД, 2001.

Этот шифр был создан в 1917 г сотрудником AT&T Corporation Гилбертом Вернамом вместе со специальным телеграфным аппаратом, который производил (на основе заданной последовательности псевдослучайных чисел) реализацию алгоритма одноразового блокнота автоматически, без участия шифровальщика.

Список литературы

Виды ГСЧ и ГПСЧ, а также методы их тестирования:

1. Кнут, Д. *Искусство программирования* / Д. Кнут – Т. 2. – М., 2002. – 788 с.
2. Харин, Ю.С. *Основы криптологии* / Ю.С. Харин, В.И. Берник, Г.В. Матвеев, С.В. Агиевич. – Минск: НОВОЕ ЗНАНИЕ, 2003. – 382 с.
3. Иванов, М.А. *Теория, применение и оценка качества генераторов псевдослучайных последовательностей* / М.А. Иванов, И.В. Чугунков. –М.: КУДИЦ-ОБРАЗ, 2003. – 240 с.

Критерии согласия:

4. Ивченко, Г.И. *Введение в математическую статистику* / Г.И. Ивченко, Ю.И. Медведев. – М.: ЛКИ, 2010. – 600 с.
5. Лагутин, М.Б. *Наглядная математическая статистика* / М.Б. Лагутин. – М.: БИНОМ, 2007. – 472 с.

Применение ГСЧ и ГПСЧ в защите информации:

6. Алферов, А.П. *Основы криптографии* / А.П. Алферов, А.Ю. Зубов, А.С. Кузьмин, А.В. Черемушкин. – М.: ГЕЛИОС АРВ, 2005. – 480 с.
7. Шнайер, Б. *Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке СИ* / Б. Шнайер. – М.: ТРИУМФ, 2003. – 816 с.

Методы Монте-Карло и задачи имитационного моделирования:

8. Лоу, А. *Имитационное моделирование* / А. Лоу, В. Келтон. – 3 изд. – СПб.: Питер, 2004. – 847 с.

Учебное издание

Ирина Сергеевна Орлова

Методы защиты информации, использующие генераторы псевдослучайных чисел

Учебное пособие

Редактор Ю.Н. Литвинова

Подписано в печать 15.08.2016. Формат 60х84/16. Бумага офсетная. Печать офсетная. Печ. л. 3,0. Тираж 300. Заказ .
Арт.–27/2016.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»
443086, Самара, Московское шоссе, 34

Издательство Самарского университета
443086, Самара, Московское шоссе, 34