Министерство образования и науки Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования ПСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт инженерных наук

Кафедра информационно-коммуникационных технологий

ЛАБОРАТОРНАЯ РАБОТА №7

ГЕНЕРАТОР СЛУЧАЙНЫХ ЧИСЕЛ. ПСЕВДОСЛУЧАЙНЫЕ ЧИСЛА

Вариант 13

по дисциплине «Моделирование»

Выполнила: Разгонова Е.В.

Группа: 0432-04

Проверил: Миронов Т.С.

Псков 2021

Задание 7.1. Генерирование псевдослучайной последовательности чисел

Задание: получите последовательности случайных чисел двумя методами: мультипликативным конгруэнтным методом (МКМ), при Q=-37, и смешанным методом. Проведите проверку качества полученных последовательностей. На основе полученных данных сравните методы получения псевдослучайных чисел.

Пояснение:

Когда речь идёт об случайных числах, то обычно понимается, что генерируемые значения не подчиняются какому-то из стандартных законов распределения, что отличает их от случайных величин.

Для имитации чисел, в основном, используется два варианта генерации – метод средних квадратов и конгруэнтный метод. Первый достаточно какой-либо общей простой, однако, описывается не формулой, последовательным алгоритмом. К сожалению этот метод неудовлетворительно: существует наличие корреляции между числами, а иногда и отсутствие характера случайности.

Самое широкое применение при моделировании на ЭВМ получили конгруэнтные методы генерации псевдослучайных последовательностей, в основе которых лежит фундаментальное понятие конгруэнтности, т.е. согласования. В математике конгруэнтность описывается следующей формулой:

$$\frac{\alpha - \beta}{k} = m$$

Трактуют её как: «Два числа α и β являются конгруэнтными, если находится такое число k, при котором деление разности этих двух чисел будет давать число m».

— Мультипликативный конгруэнтный метод

Основная формула метода:

$$x_{i+1} = ax_i \pmod{m}$$

Согласно записи, необходимо взять последнее число x_i , умножить его на коэффициент a и взять модуль полученного числа по m, то есть разделить на ax_i , и остаток считать как $x_i + 1$. Следовательно, для генерации последовательности, согласно этому методу, необходимо начальное значение x_0 , коэффициент a и m. Выбирать величины следует так, чтобы обеспечить максимальный период и минимальную корреляцию между генерируемыми числами, т.к. генерация может дать конечное множество чисел, после чего начнётся повтор.

Выбор a и x_0 может различаться в зависимости от системы, в которой моедлируется ситуация: для двоичной системы $a=8T\pm3$, в случае десятичной $-a=200T\pm Q$, в обоих случаях T — это любое целое положительное число, а велична Q может принимать одно из следующих значений \pm (3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61); x_0 — любое положительное число, которое не делится на 2 и на 5. Рассчёт числа m также зависит от системы : $m=10^d$ или $m=2^b$, где d, b — длина машинного слова.

ОВІЗПУ = 1
$$N_i$$
 = 500

Генерация псевдослучайных чисел

 $\mathbf{x}_1 := 129$ $\mathbf{m} := 10^3 = 1 \times 10^3$ $\mathbf{Q} := -37$ $\mathbf{T}_i := 10$ $\mathbf{a} := 200 \cdot \mathbf{T} + \mathbf{Q} = 1963$
 $\mathbf{i} := 1... \mathbf{N}$
 $\mathbf{x}_{i+1} := \mathbf{mod} \left(\mathbf{a} \cdot \mathbf{x}_i \cdot \mathbf{m} \right)$ Формирование псевдослучайной последовательности МКМ

 $\mathbf{x}_{i+1} = \begin{bmatrix} 227 \\ 601 \\ 763 \\ 769 \\ 547 \\ 761 \\ 843 \\ \dots \end{bmatrix}$
 \mathbf{x}_i
 \mathbf{a}_i
 \mathbf

Рисунок 1.1. Фрагмент рабочего документа MathCAD (1)

Любой генератор псевдослучайных чисел даёт лишь конечное множество чисел, после которого последовательность будет повторяться. Период, или длина последовательности L, зависит от ЭВМ и от выбранного модуля, а статистические свойства зависят от выбранного начального значения и множителя. Таким образом, выбирать a, x_0 и m следует так, чтобы обеспечить максимальный период и минимальную корреляцию между генерируемыми числами. Если правильно выбрать a и x_0 , то $L_{max} = 2^{b-2} = m/4$, когда b > 2 (двоичная система счисления) и $L_{max} = 5 * 10^{d-2} = m/20$, когда d > 2 (десятичная система счисления). Если генератор имеет слишком короткий период, то он становится непригодным.

Период генератора определяется экспериментально в ходе вычислительного эксперимента. Следует учитывать, что последовательность случайных чисел имеет апериодическую (начальную) и периодическую части. Длина апериодической части заранее неизвестна. Поэтому при определении периода генератора случайных чисел необходимо зафиксировать одно из целых чисел, полученных в результате работы процедуры генератора, и

подсчитать количество вызовов процедуры генератора до того момента, когда в результате получится то же целое число.

Методы, которые генерируют случайные числа, необходимо проверять на правильность генерируемых массивов, т.к. от этого зависит эффективность имитационной модели. Поэтому, прежде чем приступить к реализации моделирующих алгоритмов на ЭВМ, необходимо выполнить тестирование случайных чисел, которое включает в себя проверку равномерности, проверку стохастичности и проверку независимости.

Проверка равномерности выполняется по гистограмме относительных частот генерируемой случайной величины. Для ее построения интервал (0, 1) разбивается на *т* равных частей и подсчитывается относительное число попаданий значений случайной величины в каждый интервал. Чем ближе огибающая гистограммы к прямой, тем в большей степени генерируемая последовательность отвечает требованию равномерности распределения.

$$P_i = \frac{1}{m}$$

Следовательно, в каждый i-ый подинтервал попадёт N_i чисел.

Относительная частота попадания случайных чисел в каждый из подинтервалов равна $N_i/_N$. Графически она строится с помощью гистограммы, где пунктирная линия соответствует теоретическому значению P_i , а сплошная — экспериментальному $N_i/_N$.

По графику можно сделать вывод, что сгенерированная последовательность распределена равномерно. Для проверки гипотезы равномерности расчитывается критерий согласия, если она подтвердится, то случайные числа распределены равномерно.

Выдвигается следующая гипотеза: среднее значение вероятности равномерного распределения f_N^{j}/N является оценкой идеального значения f_k^{j}/N с доверительной вероятностью в 95%.

Проверка равномерности

Рисунок 1.2. Фрагмент рабочего документа MathCAD (2)

Вывод по проверке на равномерность: как видно из вычислений, гипотезу можно считать истинной, т.к. критерий согласия $\varphi = -0.273$ находится в пределах критической области $-1.96 \le \phi \le 1.96$, следовательно, проверка на равномерность пройдена.

Следующим моментом является **проверка стохастичности**. Стохастичность это также некая случайность, но проверка здесь производится нестандратным способом — процесс происходит в двоичной системе счисления.

Существует множество тестов, с помощью которых можно провести проверку. Наиболее используемыми являются интервальный тест и

автокорреляционные тесты. Интервальный тест проводится методами комбинаций и серий.

В данной работе проверка производится методом комбинаций, сущность которого сводится к определению закона распределения длин участков между единицами (нулями) или закона распределения (появления) числа единиц (нулей) в n-разрядном двоичном числе x_i . На практике длину последовательности N берут достаточно большой и проверяют все n разрядов или только l старших разрядов числа x_i .

Последовательность шагов следующая:

- 1) определение максимального числа разрядов среди всех чисел с помощью формулы $L = floor(\log(\max(x), 2));$
 - 2) перевод из десятичной системы в двоичную;
 - 3) расчёт количества единиц в числе;
 - 4) рассчёт количества чисел с определённым количеством единиц;

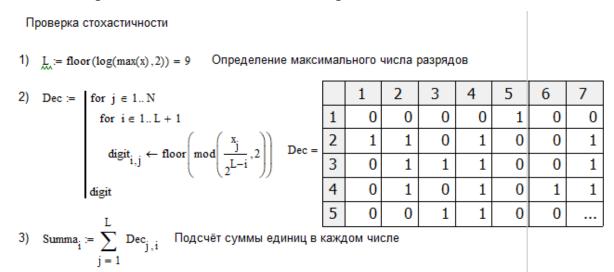


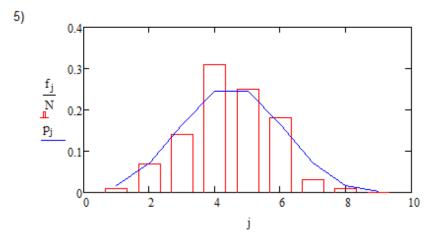
Рисунок 1.3. Фрагмент рабочего документа MathCAD (3)

5) построение гистограммы, где отображён биномиальный закон распределения, который описывает теоретический закон появления *j* единиц в l-х разрядах каждого числа. Иными словами, строится гистограмма по интервалу от минимального числа единиц до максимального с шагом равным единице.

3)
$$\operatorname{Summa}_i := \sum_{j=1}^L \operatorname{Dec}_{j,i}$$
 Подсчёт суммы единиц в каждом числе

4) SDec := sort(Summa) j := 1.. L

$$\mathbf{p}_{j} \coloneqq \frac{L!}{j! \cdot (L-j)!} \cdot 0.5^{L}$$
 Расчёт кол-ва чисел с опр. $\Delta \coloneqq 1$ $\mathbf{z}_{j+1} \coloneqq \min(\mathrm{Summa}) + \Delta \cdot \mathbf{j}$ $\mathbf{f} \coloneqq \mathrm{hist}(\mathbf{z}, \mathrm{SDec})$



finean := mean
$$\left(\frac{\mathbf{f}}{N}\right)$$
 = 0.1111 Mx := $\frac{1}{L} \cdot \sum_{j=1}^{L} p_j$ = 0.1109

$$Dx1 := \frac{1}{L} \cdot \sum_{j=1}^{L} \left(\frac{f_j}{N} - p_j \right)^2 = 7.418 \times 10^{-4}$$
 Дисперсия биномиального распределения

$$\phi 2 := \frac{\text{fmean} - Mx}{\sqrt{\frac{Dx}{L}}} = 0.094 \qquad \qquad \text{fright} := \text{qnom} \left(1 - \frac{\alpha}{2}, 0, 1\right) = 1.96$$

$$\text{fleft} := -\text{fright} = -1.96$$

Рисунок 1.4. Фрагмент рабочего документа MathCAD (4)

Гистограмма показывает количество единиц в разрядах чисел (по оси Y – экспериментальная вероятность появления ј единиц в разрядах чисел). Кривая р(j) является теоретической вероятностью появления ј единиц в разрядах числа.

Гипотеза Н0: среднее значение мат.ожидания fmean является оценкой теоретического значения мат. ожидания Mx с доверительной вероятностью 95%.

Вывод по проверке стохастичности: полученное среднее значение fmean = 0.1111 является оценкой идельного Mx = 0.1109 с вероятностью 95%. Проверку стохастичности можно считать пройденной, т.к. критерий принятия $\varphi = 0.094$ входит в доверительный интервал (-1.96, 1.96).

Последней проверкой является **проверка независимости**. Для проверки независимости берётся два массива, т.к. он основывается на корелляции. Таким образом, независимость элементов последовательности $\{x_i\}$ может быть проведена путем введения в рассмотрение последовательности $\{y_i\}$, для которых $y_i = x_{i+\tau}$, где τ — величина сдвига последовательностей. Значение сдвига также меняется в пределах некоего интервала, т.к. это номер элемента, то минимальным значением можно задать единицу до количества элементов массива минус тот шаг, с которым задаётся τ , т.е. $\tau = \overline{1, N-1}$ — диапазон допустимых значений.

В расчётной формуле коэффициента корреляции рассматриваются пары і-го элемента изначального массива и элемент массива со сдвигом. Последней парой в последовательности будет будет пара размерностью N-1 (в зависимости от заданного τ) .

$$\rho_{\tau} = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y - \bar{y})}{\sqrt{D(x) * D(y)}}$$

Данная величина не является константой для всех значений выборки, напротив, параметр считается для каждого возможного значения сдвига, т.е. получается, что ρ показывает количество возможных коэффициентов корреляции в зависимости от объёма выборки.

Далее задаётся доверительная вероятность, обычно на практике она принимает значения в интервале $\beta = 0.9 \div 0.99$. С её помощью возможно регулировать весь процесс, т.е. чем меньше берётся вероятность, тем меньше будет изменяться интервал.

При любом $\tau \neq 0$ для достаточно больших N с доверительной вероятностью β справедливо соотношение:

$$\left|\rho_{\xi\eta}(\tau)\right| \leq \beta \sqrt{\frac{1}{N}}$$

Т.к. задаётся доверительный интервал $\beta = 0.9 \div 0.99$, то теоритически с этой вероятностью все значения коэффициента корреляции должны попасть в

интервал, но на практике такая ситуация довольно жётская, поэтому, чтобы проверить количество попаданий в интервал коэффициентов корреляции, чаще всего значения сравнивают с медианой, т.е. 50% (или 0.5). Получается, что если в интервал попадает меньше 50%, то проверка провалена. Формулируется гипотеза так: «Если найденное значение $\rho_{\xi\eta}(\tau)$ находится в указанных пределах, то с вероятностью β можно утверждать, что полученная последовательность чисел x_i удовлетворяет гипотезе корреляционной независимости».

Проверка на независимость

 $\beta := 0.96$ i := 1... N $\tau := 5,10... N - 5$ Величина сдвига последовательности

$$Dxi(au) \coloneqq rac{1}{N- au} \cdot \sum_{i=1}^{N- au} \left[\left(x_i \right)^2 - rac{1}{\left(N- au
ight)^2} \cdot \left(\sum_{i=1}^{N- au} x_i
ight)^2 \right]$$
 Дисперсия входного массива

$$\mathrm{Dxt}(\tau) \coloneqq \frac{1}{\mathrm{N} - \tau} \cdot \sum_{i = 1}^{\mathrm{N} - \tau} \left[\left(\mathbf{x}_{i + \tau} \right)^2 - \frac{1}{\left(\mathrm{N} - \tau \right)^2} \cdot \left(\sum_{i = 1}^{\mathrm{N} - \tau} \mathbf{x}_{i + \tau} \right)^2 \right] \text{ Дисперсия массива со сдвигом}$$

$$\mathrm{Ppx}(\tau) := \frac{\frac{1}{\mathrm{N} - \tau} \cdot \sum_{\mathrm{i} \,=\, 1}^{\mathrm{N} - \tau} \, \left(\mathrm{x}_{\mathrm{i}} \cdot \mathrm{x}_{\mathrm{i} + \tau} \right) - \frac{1}{(\mathrm{N} - \tau)^2} \cdot \left(\sum_{\mathrm{i} \,=\, 1}^{\mathrm{N} - \tau} \, \mathrm{x}_{\mathrm{i}} \right) \cdot \sum_{\mathrm{i} \,=\, 1}^{\mathrm{N} - \tau} \, \mathrm{x}_{\mathrm{i} + \tau}}{\sqrt{\mathrm{Dxi}(\tau) \cdot \mathrm{Dxt}(\tau)}} \quad \text{Коэффициент корреляции}$$

Рисунок 1.5. Фрагмент рабочего документа MathCAD (5)

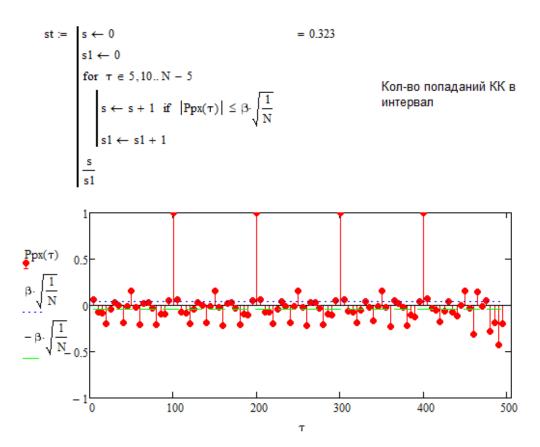


Рисунок 1.6. Фрагмент рабочего документа MathCAD (6)

На графике можно заметить, что с увеличением значения τ , уменьшается вероятность попадания корреляционного момента в доверительный интервал, а значит крепнет зависимость между числами в последовательности. Чем больше количество элементов в последовательности, тем хуже будет её качество, потому что между числами будет формироваться корреляция.

Вывод по проверке независимости: около 32.3% полученной последовательности чисел x_i находится в границах $\left(-\beta\sqrt{\frac{1}{N}},\beta\sqrt{\frac{1}{N}}\right)$, поэтому гипотеза корреляционной независимости отвергается, следовательно, проверка независимости провалена, т.к. полученный процент < 50%.

— **Смешанный метод**

В некотором роде данный метод похож на предыдущий, за исключением использования дополнительного параметра C, который строго больше 0.

$$x_{i+1} = (ax_i + C)(mod\ m)$$

С вычислительной точки зрения смешанный метод генерации последовательности позволяет уменьшить возможную корреляцию получаемых чисел, однако добавление дополнительного параметра несёт свой недостаток, например тот, что данная сумма будет действовать ровно столько раз, сколько измерений было произведено, а значит объём потребляемых ресурсов тоже будет расти.

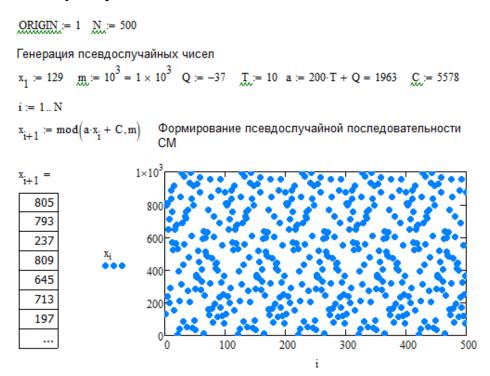


Рисунок 2.1. Фрагмент рабочего документа MathCAD (1)

Проверка равномерности

Рисунок 2.2. Фрагмент рабочего документа MathCAD (2)

Гипотеза Н0: среднее значение вероятности равномерного распределения f_{Nj}/N является оценкой идеального значения f_{Nj}/N с доверительной вероятностью в 95%.

Вывод: т.к. критерий согласия $\varphi = 0.371$ находится в пределах критической области $-1.96 \le \phi \le 1.96$, следовательно гипотеза H0 принимается и проверка на равномерность пройдена.

Проверка стохастичности Определение максимального числа разрядов 1) L := floor(log(max(x), 2)) = 9 $\begin{array}{c} \text{for } j \in 1..N \\ \text{for } i \in 1..L+1 \\ \text{digit}_{i,j} \leftarrow \text{floor} \left(\text{mod} \left(\frac{x_{j}}{2^{L-i}}, 2 \right) \right) \end{array} \right) \quad \text{Dec} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 0 \\ 4 & 0 \\ 5 & 0 \end{bmatrix}$ 1 0 1 0 1 1 1 0 3) $\operatorname{Summa}_i := \sum_{i=1}^L \operatorname{Dec}_{j,i}$ Подсчёт суммы единиц в каждом числе 4) SDec := sort(Summa) j := 1.. L $p_j := \frac{L!}{j! \cdot (L-j)!} \cdot 0.5^L \quad \text{Расчёт кол-ва чисел с опр.} \quad \underset{\text{кол-вом "1"}}{\Delta} := 1 \qquad z_{j+1} := \min(\text{Summa}) + \Delta \cdot j$ 5) 0.4 0.3 $\begin{array}{l} Dx1 := \frac{1}{L} \cdot \sum_{j = 1}^{L} \left(\frac{f_j}{N} - p_j\right)^2 = 5.802 \times 10^{-4} \quad \text{Дисперсия биномиального распределения} \\ \varphi2 := \frac{fmean - Mx}{\sqrt{\frac{Dx}{L}}} = 0.085 & fleft := -fright = -1.96 \end{array}$

Рисунок 2.3. Фрагмент рабочего документа MathCAD (3)

Гипотеза Н0: среднее значение мат.ожидания fmean является оценкой теоретического значения мат. ожидания Mx с доверительной вероятностью 95%.

Вывод: полученное среднее значение fmean = 0.1111 является оценкой идельного Mx = 0.1109 с вероятностью 95%. <u>Проверку</u> стохастичности можно считать пройденной, т.к. критерий принятия $\varphi = 0.085$ входит в доверительный интервал (-1.96, 1.96).

Проверка на независимость $\beta \coloneqq 0.96 \quad i \coloneqq 1...N \quad \tau \coloneqq 5,10...N - 5 \quad \text{Величина сдвига последовательности}$ $Dxi(\tau) \coloneqq \frac{1}{N-\tau} \cdot \sum_{i=1}^{N-\tau} \left[\left(x_i \right)^2 - \frac{1}{\left(N-\tau \right)^2} \cdot \left(\sum_{i=1}^{N-\tau} x_i \right)^2 \right] \text{ Дисперсия входного массива}$ $Dxt(\tau) \coloneqq \frac{1}{N-\tau} \cdot \sum_{i=1}^{N-\tau} \left[\left(x_{i+\tau} \right)^2 - \frac{1}{\left(N-\tau \right)^2} \cdot \left(\sum_{i=1}^{N-\tau} x_{i+\tau} \right)^2 \right] \text{ Дисперсия массива}$ co cдвигом $Ppx(\tau) \coloneqq \frac{1}{N-\tau} \cdot \sum_{i=1}^{N-\tau} \left(x_{i} \cdot x_{i+\tau} \right) - \frac{1}{\left(N-\tau \right)^2} \cdot \left(\sum_{i=1}^{N-\tau} x_i \right) \cdot \sum_{i=1}^{N-\tau} x_{i+\tau}$ Коэффициент корреляции $st \coloneqq \left[s \leftarrow 0 \right] = 0.212$ $st \leftarrow 0$ $for \quad \tau \in 5,10...N-5$ $\left[s \leftarrow s+1 \quad \text{if } |Ppx(\tau)| \le \beta \sqrt{\frac{1}{N}} \right]$ Кол-во попаданий КК в интервал

т Рисунок 2.4. Фрагмент рабочего документа MathCAD (4)

300

400

500

200

100

Гипотеза Н0: если найденное значение $\rho_{\xi\eta}(\tau)$ находится в пределах 50%, то с вероятностью $\beta=0.96$, то полученная последовательность чисел x_i удовлетворяет гипотезе корреляционной независимости.

Вывод: около 21.2 % полученной последовательности чисел x_i находится в границах $\left(-\beta\sqrt{\frac{1}{N}},\beta\sqrt{\frac{1}{N}}\right)$, поэтому гипотеза корреляционной независимости отвергается, следовательно, <u>проверка независимости провалена</u>, т.к. полученный процент < 50%. В данном случае наблюдается сильная корреляционная зависимость.

Вывод по результатам сравнения мультипликативного конгруэнтного метода и смешанного:

1) Проверка равномерности

Т.к. критерий согласия Стьюдента в мультипликативном методе $\varphi = -0.273$, а в смешанном $\varphi = 0.371$, следовательно, в смешанном методе генерируемая последовательность больше отвечает требованию равномерности распределения, чем МКМ.

2) Проверка стохастичности

Т.к. критерий согласия в мультипликативном методе $\varphi = 0.094$, а в смешанном $\varphi = 0.085$, то получается, что генерация псевдослучайных чисел в МКМ более стохастична, т.е. более случайна.

3) Проверка независимости

При проценте попадания в 96% интервал, мульпликативный метод показывает около 32.3 % случайных чисел из полученной последовательности, смешанный метод — 21.2 %. Значит проверку на независимость МКМ прошёл лучше, чем СМ, в котором наблюдается более явная корреляция.*

*Примечание: следует понимать, что при каждом изменении x_i переменной происходит эквивалентное изменение $y_{i+\tau}$ переменной в том же направлении. На графиках обоих способов генерации присутствует несколько точек, где значение коэффициента корреляции равно 1. Это связано с дублированием, т.е. это та же последовательность, но генерируемая несколько раз в связи с наличием периодичности в генераторе.