Получение на ЭВМ равномерно распределенных псевдослучайных чисел

Цель работы – изучение методов получения на ЭВМ равномерно распределенных псевдослучайных чисел и тестов проверки их качества.

Теоретические сведения

При моделировании систем на ЭВМ программная имитация случайных воздействий любой сложности сводится к генерированию некоторых стандартных (базовых) процессов и к их последующему функциональному преобразованию. Таким базовым процессом является последовательность чисел $\{x_i\}=x_0,\ x_1,\ \ldots,\ x_N$, представляющих собой реализации независимых, равномерно распределенных на интервале (0, 1) случайных величин $\{\varepsilon_i\}=\varepsilon_0,\ \varepsilon_1,\ \ldots,\ \varepsilon_N$. Но на ЭВМ невозможно получить идеальную последовательность случайных чисел потому, что на ней можно оперировать только с конечным множеством чисел. Кроме того, для получения значений x случайной величины ε используются формулы (алгоритмы). Поэтому такие последовательности, являющиеся по своей сути детерминированными, называются псевдослучайными.

Наибольшее применение в практике моделирования на ЭВМ для генерации последовательностей псевдослучайных чисел находят алгоритмы вида

$$x_{i+1} = \Phi(x_i), \tag{1}$$

представляющие собой рекуррентные соотношения первого порядка, для которых начальное число x_0 и постоянные параметры заданы.

Рассмотрим некоторые процедуры получения последовательностей равномерно распределенных псевдослучайных чисел.

Метод серединных квадратов

Пусть имеется $2n$-разрядное число, меньшее 1:

$$x_i = 0,a_1,a_2,....,a_{2n}$$

1. Возведем его в квадрат:

$$x_i^2 = 0,b_1,b_2,...,b_{4n}$$

2. Отберем средние $2n$ разрядов $x_{i+1} = 0,b_{n+1},b_{n+2},...,b_{3n}$ которые будут являться очередным числом псевдослучайной последовательности.

Пример, если начальное число x_0=0,2152, то $(x_0)^2$=0,04631104. Тогда x_1=0,6311, затем $(x_1)^2$=0,39828721, тогда x_2=0,8287, и т. д.

Недостаток метода:

Наличие корреляции между числами последовательности, в некоторых случаях может отсутствовать.

Метод середины произведения

Метод является модификацией метода серединных квадратов и состоит в том, что два $2n$-значных числа перемножаются и средние $2n$

цифр этого произведения принимаются в качестве следующего числа последовательности. Таким образом, если

$$x_{i-1} = a_1, a_2, ..., a_{2n},$$
$$x_i = 0, b_1, b_2, ..., b_{2n},$$

то для получения числа x_{i+1} необходимо перемножить x_{i-1} и x_i:

$$x_{i+1} = 0, c_1, c_2, ..., c_{4n}.$$

Отберем средние $2n$ цифр этого произведения

$$x_{i+1} = 0, c_{n+1}, c_{n+2}, ..., c_{3n}.$$

Несмотря на то, что данный метод имеет тенденцию к вырождению, он обеспечивает лучшее качество псевдослучайных чисел, чем у чисел, полученных с помощью метода серединных квадратов.

Конгруэнтный метод

Широкое применение для получения последовательностей псевдослучайных равномерно распределенных чисел получили конгруэнтные процедуры генерации, которые могут быть реализованы мультипликативным либо смешанным методом. Конгруэнтные процедуры являются чисто детерминированными, т.к. описываются в виде рекуррентного соотношения, когда функция (1) имеет вид:

$$x_{i+1} = \lambda x_i + \mu \pmod M, \quad (2)$$

где x_i, λ, μ, M – неотрицательные целые числа.

Раскрывая (2) получим

$$x_{i+1} = \lambda^i x_0 + (\lambda^i - 1)\mu/(\lambda - 1) \pmod M). \quad (3)$$

Если задано начальное значение x_0, множитель λ и аддитивная константа μ, то (3) однозначно определяет последовательность целых чисел $\{x_i\}$, составленную из остатков от деления на M, членов последовательности

$$\{\lambda^i x_0 + (\lambda^i - 1)\mu/(\lambda - 1)\}$$

Таким образом, для любого $i \geq 1$ справедливо неравенство $x_i < M$. По целым числам последовательности $\{x_i\}$ можно построить последовательность $\{x_i^*\} = \{x_i / M\}$ рациональных чисел из единичного интервала (0, 1).

Мультипликативный метод

Задает последовательность неотрицательных целых чисел $\{x_i\}$, не превосходящих M, по формуле

$$x_i = \lambda x_i \pmod M,$$

т.е. это частный случай (2) при $\mu = 0$.

Пример. Если M=113, λ=101 и x_0=50, то

x_0^*=50/113 = 0,4424,

x_1=101·50 (mod 113) = 78,

x_1^*=78/113 = 0,6902 и т.д.

Методы проверки качества псевдослучайных чисел с равномерным законом распределения

Методы (в дальнейшем, тесты) проверки качества псевдослучайных чисел делятся на три группы:

а) тесты проверки ″случайности″ последовательности псевдослучайных чисел;

б) тесты проверки равномерности закона распределения;

в) тесты проверки независимости последовательности.

Первые два теста основываются на статистических критериях согласия, из которых наиболее употребительным является статистический критерий согласия χ^2 (Пирсона).

Пусть имеется η – случайная величина, о законе распределения которой выдвигается некоторая гипотеза, X – множество возможных значений η. Разобьем X на m попарно непересекающихся множеств $X_1, X_2, \ldots, X_m$, таких, что

$$P\{\eta \in X_j\} = p_j > 0 \text{ при } j = 1, 2, \ldots, m,$$

$$p_1 + p_2 + \ldots + p_m = P\{\eta \in X\} = 1.$$

Выберем N независимых значений $\eta_1, \eta_2, \ldots, \eta_N$ и обозначим через ν_j количество значений $\eta \in X_j$. Очевидно, что математическое ожидание ν_j равно Np_j, т.е. $M(\nu_j) = Np_j$.

В качестве меры отклонения всех ν_j от Np_j выбирается величина

$$\chi_N^2 = \sum_{j=1}^{m} \frac{\left(\nu_j - Np_j\right)^2}{Np_j}.$$

При достаточно большом N величина χ_N^2 хорошо подчиняется закону распределения χ^2 с $(m-1)$ степенью свободы:

$$P\{\chi_N^2 < x\} = \int_0^x k_{m-1}(x)\,dx, \qquad (4)$$

где $k_{m-1}(x)$ – плотность распределения χ^2 с $(m-1)$ степенью свободы.

С помощью формулы (4) при заданном уровне значимости β (обычно β=0.95) можно определить нижнюю χ_H^2 и верхнюю χ_B^2 границы области возможного принятия гипотезы (доверительного интервала). Для этого нужно решить соответственно следующие уравнения:

$$P\{\chi_N^2 > \chi_H^2\} = \int_{\chi_H^2}^{\infty} k_r(x)\,dx = \beta,$$

$$P\{\chi_N^2 > \chi_B^2\} = \int_{\chi_B^2}^{\infty} k_r(x)dx = \gamma,$$

где $\gamma = 1 - \beta$, $r = m - 1$.

В приложении 1 имеется таблица, в которой приведены решения уравнения

$$P\{\chi_N^2 > x\} = p,$$

где $x = \chi_H^2$ или $x = \chi_B^2$, $p = \beta$ или $p = \gamma$.

Тесты проверки "случайности"

На практике обычно применяют два теста проверки "случайности": тест проверки серий и тест проверки частот и пар.

Тест проверки серий предусматривает разбиение случайных цифр в исследуемой последовательности на элементы двух родов – первого и второго.

Серией называется любой отрезок последовательности цифр, состоящий из следующих друг за другом элементов одного и того же рода. Например, если в последовательности цифр

$\boxed{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k}$	$\boxed{\varepsilon_{k+1}, \varepsilon_{k+2}, \dots, \varepsilon_{k+l}}$	$\boxed{\varepsilon_{k+l+1}, \varepsilon_{k+l+2}, \dots, \varepsilon_s}$
↓	↓	↓
серия 1-го рода длины k	серия 2-го рода длины l	серия 1-го рода длины $s - k - l$

$\varepsilon_1 \neq \varepsilon_2 \neq \dots \neq \varepsilon_k \neq \varepsilon_{k+1}$, $\varepsilon_{k+1} = \varepsilon_{k+2} = \dots = \varepsilon_{k+l}$ и $\varepsilon_{k+l} \neq \varepsilon_{k+l+1} \neq \dots \neq \varepsilon_s$, то цифры $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$ образуют серию первого рода длины k, цифры $\varepsilon_{k+1}, \varepsilon_{k+2}, \dots, \varepsilon_{k+l}$ образуют серию второго рода длины l и цифры $\varepsilon_{k+l+1}, \varepsilon_{k+l+2}, \dots, \varepsilon_s$ также образуют серию первого рода длины $s-k-l$. Иногда для удобства элементы серий первого рода обозначают знаками "–" (минус), а второго рода – знаками "+" (плюс). В этом случае рассматриваемая последовательность будет иметь такой вид:

$\boxed{- - \dots -}$	$\boxed{+ + \dots +}$	$\boxed{- - \dots -}$
↓	↓	↓
k минусов	l плюсов	$s-k-l$ минусов

Подсчитаем количество z_l серий второго рода длины l в последовательности псевдослучайных цифр $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$. Пусть $l=1, 2, \dots, m$ и z'_{m+1} – количество серий второго рода с $l \geq m + 1$ (они объединяются в одну группу). Обозначим общее количество серий через

$$z = z_1 + z_2 + \dots + z_m + z'_{m+1}.$$

Величина χ_z^2 с m степенями свободы вычисляется по формуле:

$$\chi_z^2 = \sum_{l=1}^{m} \frac{\left(z_l - p_l z\right)^2}{p_l z} + \frac{\left(z'_{m+1} - p'_{m+1} z\right)^2}{p'_{m+1} z},$$

где $p_l = 9 \cdot 10^{-l}, p'_{m+1} = 10^{-m}$.

Если, с заданным уровнем значимости β, значение χ_z^2 попадает в доверительный интервал (χ_H^2, χ_B^2), то тест проверки серий удовлетворяется.

В практике встречается также другая разновидность теста проверки серий, когда к элементам серий первого рода относят цифры, меньшие 0,5, а к элементам серий второго рода – не меньшие 0,5.

При достаточно большом объеме выборки $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ (практически при $N \geq 20$) и уровне значимости $\beta = 0.95$ нижний предел $z^н$ общего числа серий равен:

$$z^н = \frac{1}{2}\left(N + 1 - 1.65\sqrt{N-1}\right),$$

а нижний предел числа серий элементов первого $z^н_{п.р.}$ и второго $z^н_{в.р.}$ родов равен:

$$z^н_{п.р.} = z^н_{в.р.} = \frac{1}{4}\left(N - 1.65\sqrt{N+1}\right).$$

Максимальная длина серий не должна быть больше, чем

$$l_{\max} = 3.3(\lg N + 1).$$

Тест проверки равномерности закона распределения

Данный тест строится на основе применения критерия согласия χ^2. Пусть имеется выборка $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ псевдослучайных чисел в интервале (0,1). Интервал (0,1) изменения случайной величины ε разбивается на m интервалов x_j, $j = 1, 2, \ldots, m$, очевидно, что $x_m = 1$, а нижняя граница первого интервала равна нулю. Обычно принимают m=10÷20.

Далее производится определение вероятности p_j попадания случайной величины ε в j-й интервал. Для равномерного на интервале (0,1) закона распределения $p_j = x_j - x_{j-1}$. Затем определяется величина ν_j, j=1,2,…,m – число попаданий случайной величины ε в j-й интервал и подсчитывается величина

$$\chi_N^2 = \sum_{j=1}^{m} \frac{\left(\nu_j - p_j N\right)^2}{p_j N},$$

распределенная по закону χ^2 с (m–1) степенью свободы. По заданному уровню значимости β путем решения уравнений

$$P\ \{\ \chi_N^2 > x\ \} = p,$$

где $x = \chi_H^2$ или $x = \chi_B^2$, $p = \beta$ или $p = \gamma$, (с помощью таблицы, приведенной в приложении 1) можно определить нижнюю χ_H^2 и верхнюю χ_B^2 границы доверительного интервала. Если подсчитанное значение χ_N^2 не попадает в доверительный интервал (χ_H^2, χ_B^2), то гипотезу о равномерном законе распределения случайной величины ε следует отвергнуть.

Дополнительно можно подсчитать эмпирическое математическое ожидание

$$\overline{\varepsilon} = \frac{\sum_{j=1}^{N} \varepsilon_j}{N}$$

и эмпирическую дисперсию

$$S^2 = \frac{1}{N-1} \sum_{j=1}^{N} \left(\varepsilon_j - \overline{\varepsilon}\right)^2$$

и сравнить их с теоретическими значениями соответственно 0,5 и 1/12. Если $\overline{\varepsilon}$ > 0,5 и S^2 > 1/12, то гипотеза отвергается.

Для математического ожидания можно для заданного уровня значимости β необходимо определить также доверительный интервал:

$$0{,}5 - \delta \le \overline{\varepsilon} \le 0{,}5 + \delta,$$

где δ определяется из уравнения:

$$2\Phi\left(\sqrt{12}\delta\sqrt{N}\right) = \beta.$$

Значения интеграла вероятностей Φ(x) приведены в приложении 2.

Полезно бывает сравнить также теоретическую функцию распределения $F(x)$ и теоретическую плотность распределения $f(x)$ случайной величины ε с экспериментально полученными функцией распределения $F^*(x)$ и гистограммой частот.

Известно, что для случайной величины, равномерно распределенной на интервале (0,1):

$$F(x)=\begin{cases}0,\ при\ \ x<0,\\ x, при\ \ 0\le x\le 1,\\ 1, при\, x>1.\end{cases}$$

$$f(x)=\begin{cases}0,\ при\ \ x<0,\\ 1, при\ \ 0\le x\le 1,\\ 0,\ при\, x>1.\end{cases}$$

По известной выборке из N значений случайной величины ε экспериментальная функция распределения $F^*(x)$ определяется следующим образом:

$$F^*(x)=\frac{S_N(x)}{N},$$

где $S_N(x)$ равно количеству значений ε < x.

Гистограмма частот, являющаяся аналогом плотности распределения, строится следующим образом. Весь интервал ($x_{\min}$, $x_{\max}$) от наименьшего значения $x_{\min}$ до наибольшего значения $x_{\max}$ полученной выборки случайной величины разбивается на q равных промежутков длины h. Затем определяется число значений ν_i выборки, попавших в i-ый промежуток, после чего для каждого $1 \le i \le q$ строится прямоугольник с основанием на i-ом промежутке и высотой ν_i / h. Полученный чертеж называется гистограммой частот или просто гистограммой. Отметим, что при таком построении площадь i-го прямоугольника равна $h\cdot(\nu_i/h) = \nu_i$, т.е. числу значений случайной величины, попавших в i-ый промежуток, а площадь всей гистограммы равна объему выборки.

Тесты проверки независимости последовательности псевдослучайных чисел

В основе этих методов лежит представление полученных псевдослучайных чисел в качестве реализации дискретного стационарного случайного процесса $x(t)$.

Для количественной оценки степени некоррелированности последовательности псевдослучайных чисел ε_1, ε_2, …,ε_N применяется способ, заключающийся в определении коэффициента корреляции $\rho(\varepsilon_{i,i})$ между элементом ε_i последовательности и его номером i:

$$\rho\left(\varepsilon_{i,i}\right)=\frac{\frac{1}{N}\sum_{i=1}^{N} i\cdot\varepsilon_i-\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\cdot\frac{N+1}{2}}{\sqrt{\left[\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2-\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\right)^2\right]\frac{N^2-1}{12}}}.$$

Если при заданном уровне значимости β

$$\left|\rho\left(\varepsilon_{i,i}\right)\right| > \rho_{\max} = z_{\beta}\frac{1-\rho^{2}\left(\varepsilon_{i,i}\right)}{\sqrt{N}},$$

где $\rho_{\max}$ – верхняя граница доверительного интервала, а z_{β} определяется из уравнения:

$$2\Phi(z_{\beta}) = \beta,$$

то считается, что имеет место корреляционная связь между псевдослучайными числами. В противном случае можно принять гипотезу об их независимости.

Порядок выполнения работы

1. Изучить методы получения на ЭВМ равномерно распределенных псевдослучайных чисел и тесты проверки их качества.

2. Составить программы получения на ЭВМ равномерно распределенных псевдослучайных чисел, состоящую из 300 чисел, методами: серединных квадратов, середины произведений, мультипликативным и оценить качество полученной мультипликативным методом последовательности тремя тестами - случайности, равномерности и независимости.

Содержание отчета

Отчет должен содержать:

1. Таблицы полученных псевдослучайных чисел.

2. Результаты проверки качества полученных псевдослучайных чисел с представлением:

а) при проверке на "случайность" – расчетов величин χ_z^2 и их доверительных интервалов, а также величин $z^{н}, z^{н}_{п.р.}, z^{н}_{в.р.}$ и $l_{\max}$;

б) при проверке на равномерность закона распределения – расчетов эмпирических значений математического ожидания, дисперсии псевдослучайной величины ε, доверительных интервалов для эмпирического математического ожидания, величин χ_N^2 и их доверительных интервалов, а также графиков теоретической и экспериментальной функции распределения, теоретической плотности распределения и полученной гистограммы;

в) при проверке на независимость – расчетов коэффициента корреляции $\rho(\varepsilon_{i,i})$ и верхней границы $\rho_{\max}$ его доверительного интервала.

3. Формулы для расчета.

Приложение 1

Таблица решений уравнения $P\{\chi_N^2 > x\} = q$ для распределения χ^2 с r степенями свободы

β	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30
1	0.0002	0.0006	0.0039	0.016	0.064	0.148	0.455	1.07
2	0.020	0.040	0.103	0.211	0.446	0.713	1.386	2.41
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.66
4	0.30	0.43	0.71	1.06	1.65	2.19	3.36	4.9
5	0.55	0.75	1.14	1.61	2.34	3.00	4.35	6.1
6	0.87	1.13	1.63	2.20	3.07	3.83	5.35	7.2
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.4
8	1.65	2.03	2.73	3.49	4.59	5.63	7.34	9.5
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.7
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.8
11	3.1	3.6	4.6	5.6	7.0	8.1	10.3	12.9
12	3.6	4.2	5.2	6.3	7.8	9.0	11.3	14.0
13	4.1	4.8	5.9	7.0	8.6	9.9	12.3	15.1
14	4.7	5.4	6.6	7.8	9.5	10.8	13.3	16.2
15	5.2	6.0	7.3	8.5	10.3	11.7	14.3	17.3
16	5.8	6.6	8.0	9.3	11.2	12.6	15.3	18.4
17	6.4	7.3	8.7	10.1	12.0	13.5	16.3	19.5
18	7.0	7.9	9.4	10.9	12.9	14.4	17.3	20.6
19	7.6	8.6	10.1	11.7	13.7	15.4	18.3	21.7
20	8.3	9.2	10.9	12.4	14.6	16.3	19.3	22.8

γ	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1	1.64	2.7	3.8	5.4	6.6	7.9	9.5	10.83
2	3.22	4.6	6.0	7.8	9.2	11.6	12.4	13.8
3	4.64	6.3	7.8	9.8	11.3	12.8	14.8	16.3
4	6.0	7.8	9.5	11.7	13.3	14.9	16.9	18.5
5	7.3	9.2	11.1	13.4	15.1	16.3	18.9	20.5
6	8.6	10.6	12.6	15.0	16.8	18.6	20.7	22.5
7	9.8	12.0	14.1	16.6	18.5	20.3	22.6	24.3
8	11.8	13.4	15.5	18.2	20.1	21.9	24.3	26.1
9	12.2	14.7	16.9	19.7	21.7	23.6	26.1	27.9
10	13.4	16.0	18.3	21.2	23.2	25.2	27.7	29.6
11	14.6	17.3	19.7	22.6	24.7	26.8	29.4	31.3
12	15.8	18.5	21.0	24.1	26.2	28.3	31	32.9
13	17.0	19.8	22.4	25.5	27.7	29.8	32.5	34.5
14	18.2	21.1	23.7	26.9	29.1	31	34	36.1
15	19.3	22.3	25.0	28.3	30.6	32.5	35.5	37.7
16	20.5	23.5	26.3	29.6	32.0	34	37	39.2
17	21.6	24.8	27.6	31.0	33.4	35.5	38.5	40.8
18	22.8	26.0	28.9	32.3	34.8	37	40	42.3
19	23.9	27.2	30.1	33.7	36.2	38.5	41.5	43.8
20	25.0	28.4	31.4	35.0	37.6	40	43	45.3

Приложение 2

Таблица значений интеграла вероятностей $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.0000	1.20	0.3849	2.40	0.4918
0.05	0.0199	1.25	0.3944	2.45	0.4929
0.10	0.0398	1.30	0.4032	2.50	0.4938
0.15	0.0596	1.35	0.4115	2.55	0.4946
0.20	0.0793	1.40	0.4192	2.60	0.4953
0.25	0.0987	1.45	0.4265	2.65	0.4960
0.30	0.1179	1.50	0.4332	2.70	0.4965
0.35	0.1368	1.55	0.4394	2.75	0.4970
0.40	0.1554	1.60	0.4452	2.80	0.4974
0.45	0.1736	1.65	0.4505	2.85	0.4978
0.50	0.1915	1.70	0.4554	2.90	0.4981
0.55	0.2088	1.75	0.4599	2.95	0.4984
0.60	0.2257	1.80	0.4641	3.00	0.4987
0.65	0.2422	1.85	0.4678	3.05	0.4989
0.70	0.2580	1.90	0.4713	3.10	0.4990
0.75	0.2734	1.95	0.4744	3.15	0.4992
0.80	0.2881	2.00	0.4773	3.20	0.4993
0.85	0.3023	2.05	0.4798	3.25	0.4994
0.90	0.3159	2.10	0.4821	3.30	0.4995
0.95	0.3289	2.15	0.4842	3.35	0.4995
1.00	0.3413	2.20	0.4861	3.40	0.4996
1.05	0.3531	2.25	0.4878	3.45	0.4997
1.10	0.3643	2.30	0.4893	3.50	0.4998
1.15	0.3749	2.35	0.4906	3.75	0.4999